Final Report

Contract N00014-91-J-4159

Improving problem-solving and decision-making skills under stress: Prediction and training

Phillip L. Ackerman
Ruth Kanfer

University of Minnesota

The research summarized in this report was sponsored by the Psychological Sciences Division, Office of Naval Research, under Contract No. N00014-91-J-4159, Contract Authority Identification Number, NR 4422-543.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)

2. REPORT DATE
10 September 94

3. REPORT TYPE AND DATES COVERED
Final Report (9/91 - 7/94)

4. TITLE AND SUBTITLE
Improving problem-solving and decision-making skills under stress: Prediction and training

5. FUNDING NUMBERS
C N00014-91-J-4159

6. AUTHOR(S)
Phillip L. Ackerman
Ruth Kanfer

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Minnesota, Department of Psychology
N218 Elliott Hall, 75 E. River Rd.
Minneapolis, MN 55455

8. PERFORMING ORGANIZATION REPORT NUMBER
Report 94-01

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
Cognitive Sciences Program (Code 1142CS)
800 N. Quincy Street
Arlington, VA 22217-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified/Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

This investigation draws on a theoretical framework developed by Kanfer & Ackerman to: (a) assess individual differences in reactivity to stress, particularly from information processing, and (b) to develop a battery of cognitive ability tests/self-regulation measures for predicting performance under stress. Three studies were conducted. Two studies used a dynamic Target/Threat Identification Task under time-pressure. The remaining study used a complex air traffic controller task (TRACON) over extended practice. Personality, motivation, and self-regulatory activity measures are often significantly associated with performance under task load. However, the mixture of cognitive ability influences and stress-reactions associated with performance on cognitive and perceptual speed abilities account for the major amount of individual differences in performance, and also account for variance that is common to the non-ability predictors and task performance. The Dial Reading and Directional Headings Tests are promising predictors of performance in stressful information processing activities. We conclude that such measures have definite advantages to self-reports in predicting individual differences in performance under stress.

14. SUBJECT TERMS
Individual differences, cognitive abilities, skills, stress, self-regulation

15. NUMBER OF PAGES
79

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

I. Abstract

The current investigation draws on a theoretical framework developed by Kanfer & Ackerman in order to (a) assess individual differences in reactivity to induced stress, particularly in the domain of information processing components of performance, and (b) to develop a battery of cognitive ability tests/self-regulation measures than can be used to predict individual differences in performance under conditions of stress, both at a local level (i.e., specific task performance) and at a global level (e.g., overall job performance). Two laboratory studies and one field study were conducted to converge on these issues. The first laboratory study and the field study used a dynamic Target/Threat Identification Task under time-pressure (in Study 1, information processing stress was induced by the addition of a secondary auditory memory task). The second laboratory study used a complex air traffic controller simulation task (TRACON), over extended task practice. Results of the studies indicate that measures of personality, motivation, and self-regulatory activity are often significantly associated with performance under task load conditions. However, the mixture of cognitive ability influences and stress-reactions associated with performance on standardized measures of cognitive and perceptual speed abilities not only account for the major amount of individual differences in performance, but also account for variance that is common to the non-ability predictors and task performance. Two tests, which first appeared to represent complex perceptual speed ability, the Dial Reading Test and the Directional Headings Test, appear to be especially promising predictors of individual differences in performance in stressful information processing activities. We conclude that such measures have definite advantages to self-report measures in predicting individual differences in performance under stress.

Table of Contents

II. Project Overview

Group decision-making (whether hierarchical or consensus-based) under time-stress with incomplete information is partly a function of the interaction of team members and partly a function of individual competencies and resistance to stress. Although safeguards may be instituted for identification of individual performance deficits under stressful conditions, a fundamental problem for group decision-making is that performance is critically limited by individual competencies. **The problem is that the chain (of decision-making) breaks at the weakest link (the individual whose performance significantly degrades under stress).** When faced with stress, performance of tasks requiring hypothesis generation and confirmation, option selection, target acquisition and identification, can be impaired by three major types of adverse individual reactions. These reactions are: (1) cognitive (problem-solving and decision-making), (2) motivational (effort and attention regulation), and (3) affective (emotional reactivity).

Although there has been a great deal of research that focuses on the long-term and short-term effects of stress on human performance, interactions between individual capacities and limitations and stress have been under-researched, and as such, are still poorly understood. Nonetheless, research conducted in this area, and in the broader domain of differential psychology indicates that performance of individuals in isolation, and individuals within teams or work-groups under stressful conditions may be predicted from measures administered prior to exposure to stress. In addition, other research has indicated that interventions (especially in the domain of skill training for self-regulatory processes), may partially 'immunize' individuals from some stress-related deficits in information processing.

By focusing on the individual as the unit of analysis, generalizations for selection, classification, and training applications will be possible. Although the work-group might be expected to be more than the sum of its individual members (e.g., command information center (CIC) situations with many individuals in the work group), the current approach focuses on avoiding potential problems by identifying (and potentially setting the stage for remediation of) those individuals who will be the weakest links in the chain-of-command.

This project focuses on identification and prediction of individual differences in cognitive performance under stressful conditions. In addition we attempt to integrate the prediction of individual differences with other work relating to remediation and training.

1. The main task to be accomplished by this project is to develop a set of tests that will allow prediction of individuals who, although they may adequately function under normal conditions, will fail in performance during stressful conditions. (This part of the project was predicated on work that we have performed in identification of individual differences in future performance of air traffic controllers -- a non-combat job that has similar demands to Navy-relevant tasks decision-making under time stress, where many lives hang in the balance of correct decision making, and vigilance in attention is paramount. (Ackerman, 1992, Ackerman & Kanfer, 1993).

2. The secondary task for this project is to identify those individuals who may be identified as "trainable" in resistance to adverse reactions under stressful conditions, such that they may be subjected to a remediation strategy -- that will immunize their problem-solving/decision-making skills from stress. This work is an extension of a project we completed for the U.S. Air Force (Kanfer & Ackerman, 1990). In that project, we trained U.S.A.F. recruits to maintain motivational control (attention and effort) and emotion control (negative emotions) during performance of a demanding learning task. The training aspect of the current project will focus on preserving individual competencies under stress.

The current investigation thus seeks to: (a) assess individual differences in reactivity to induced stress, particularly in the domain of information processing components of performance, and (b) to develop a battery of tests/self-report measures than can be used to predict individual differences in stress reactivity, both at a local level (i.e., specific task performance) and at a global level (e.g., overall job performance, attrition, etc.). A series of field and lab-based experiments were completed, in order to develop, validate, and refine the battery of tests/self-report measures. The field-based experiment focuses on integrating our approach to measurement and prediction of individual differences in stress-related information processing deficits. The lab-based experiments focus on stress-related deficits that take place under complex skill acquisition and skill maintenance conditions. These studies are expected to yield a test/task battery that will serve to predict (prospective) and evaluate (concurrent) individual differences in stress-induced changes to cognitive information processing and self-regulatory processing.

III. Background

A. Review of skill acquisition principles

The starting point for a consideration of ability-performance relations regarding skills is not with individual differences, but rather with general theories of skill acquisition. Fortunately, at the level of analysis needed for this discussion, most current theories of perceptual-motor learning are in agreement about the nature of the learning process. Although learning is typically thought of as a continuous process, without breaks or hierarchical plateaus (e.g. see Newell and Rosenbloom. 1981), theorists have often settled on a description of learning that is broken into phases or stages of skill acquisition. More details can be found in the respective descriptions of the theories, but for the current level of analysis, there is substantial similarity in the descriptions of three major phases of skill acquisition, as they are described by Fitts and Posner (1967), Anderson (1982, 1985), and Shiffrin and Schneider (1977).

Phase 1 -- Declarative Knowledge. The first phase of skill acquisition starts with the learner's initial confrontation with the task. During this phase, the learner begins to understand the basic task requirements, the rules for task engagement. In other words, the learner formulates a general idea of what is required of him or her. The term "declarative knowledge" refers to "knowledge about facts and things" (Anderson, 1985, p. 199). For a word processing operator, information that would have to be encoded as declarative knowledge would presumably include rules of starting the program, organization of commands (e.g. menu structure), effects of command implementation on the text (e.g. understanding the difference between "single occurrence" and "global" search and replace commands), location of command keys, and the use of macro (or multiple) commands. The essence of declarative knowledge is that it is represented in a form that generally allows conscious-mediated retrieval. Thus, a test of knowledge at this stage of skill acquisition often takes the form of a written examination of facts and procedures. As the research of Schneider et al. has pointed out, performance at this phase of skill acquisition is slow, attentionally effortful, and error prone (e.g., see Fisk, Ackerman, & Schneider, 1987).

Phase 2 -- Knowledge Compilation. Following the development of the necessary declarative knowledge base that is required to move beyond a trial-and-error task engagement, the learner can proceed to the next phase of skill acquisition. This phase has been termed the "associative" stage of skill acquisition by Fitts and Posner (1967) because the key element of performance improvements that take place during this phase come about through increasing the strength and efficiency of associations between stimulus conditions and appropriate response patterns. In this sense, the "associations" are formed in the first phase of skill acquisition, but are strengthened in the second phase of skill acquisition. Anderson has coined the term "knowledge compilation" from a computer programming analogy. That is, performance at the first phase of skill acquisition can be thought of as writing a computer program interactively. As each line of code is written, the system indicates whether the code is syntactically correct or incorrect. However, at the "compilation" phase, the program has been completed and is compiled to be run more quickly and efficiently. Regardless of the analogy adopted, this phase of skill acquisition is typically marked by substantial increases in the speed of task accomplishment, with a concomitant increase in the accuracy of performance. While conscious mediation often takes place during the early part of this phase, with additional practice, the conscious mediation often becomes an epiphenomenon -- it no longer plays a role in determining the speed and accuracy of performance.

Phase 3 -- Procedural Knowledge. When the learner has reached a level of skill such that performance is characterized as requiring minimal attentional effort, but at the same time is fast and accurate, the knowledge required to perform the task has become "automatized" (Shiffrin and Schneider, 1977). Anderson chose the term "proceduralized knowledge" to refer to "knowledge about how to perform various cognitive activities" (Anderson, 1985; p. 199). In contrast to declarative knowledge, proceduralized knowledge does not require conscious mediation. In fact, if the task is sufficiently proceduralized, so that declarative knowledge is no longer involved in task accomplishment, declarative knowledge of "how" the task is performed can be forgotten, with no decrement to task performance (Schneider & Fisk, 1982). Numerous common examples of such situations can be found, from the person who can effortlessly dial a familiar phone number sequence (but has difficulty writing down the actual number), to a person who operates a complex piece of machinery (but has difficulty in reporting what it is that he/she does to a knowledge engineer).

B. Review of structure of cognitive/intellectual abilities

In order to discuss individual differences in skill acquisition, and particularly the role of abilities in predicting performance at the three phases of skill acquisition, it is necessary to explicate a structure of cognitive/intellectual abilities. Although the literature contains numerous competing theories of the structure of human intellectual abilities (e.g. Guilford, 1982; Horn and Cattell, 1966; Vernon, 1961), there is a familial similarity among most current perspectives. These theories have been termed "hierarchical."

Explicit in most of these theories is a general intellectual ability factor (or *g*). The *g* ability factor represents the highest node in a hierarchy of ability factors. The influence of such a factor has been estimated by Vernon as reflecting anywhere from roughly 20-40% of the variance in a population of "all human abilities." However, the theories diverge when it comes to identification of factors that constitute the nodes below *g*. However, all theories appear to be in agreement about the nature of the hierarchy. That is, the general factor represents the broadest ability, and factors at the next level represent broad, or major group factors (e.g. Verbal:Educational, Practical:Mechanical, as in Vernon's theory). Each of the abilities at this broad group factor node may be further fragmented to reveal their constituent abilities. For example, at the next node, the Verbal ability factor might fragment into Vocabulary, Reading Comprehension, Associational Fluency, and so forth. These lower ability nodes may, in turn, be further subdivided to allow representation of the different test formats for assessing the specific abilities, and so on.

A particularly useful heuristic approach to describing the structure of abilities that is also consistent with the hierarchical approach is one developed by Marshalek, Lohman and Snow (1983; see also Snow, Kyllonen, and Marshalek, 1984) called the radex. The graphic representation of the structure of abilities is with a circle, as shown in Figure 1.

Two salient characteristics of this structure should be noted. First, the proximity of a test or measure to the center of the circle is determined by the "complexity" of the material being tested. As complexity of material increases, the measure shares more in common with *g* or with tests of general intellectual abilities (e.g. reasoning). As the complexity of test material decreases, the measure has less in common with *g*, and indeed, more in common with narrower abilities and skills. The second salient characteristic of this model is that "content" abilities represent different segments, or slices of the circle. This particular aspect of the structure illustrates the nature of "group" factors of spatial, verbal, numerical (and, perhaps, mechanical) abilities. That is, tests that share the same or similar contents, also have an increased degree of common variance. Thus, the location of a particular test in this structure is a function of the test's complexity of required information processing *and* the type of content upon which the test is constructed.

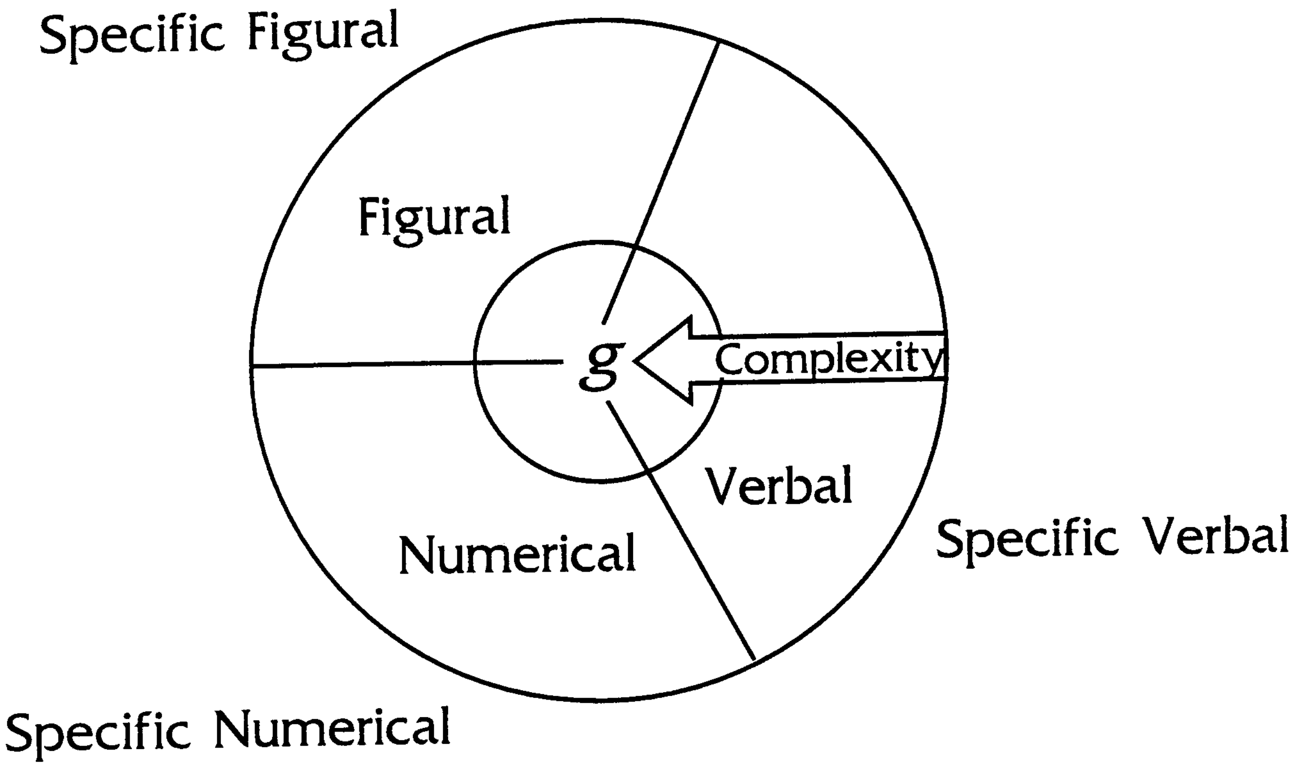


Figure 1. Structure of intelligence map proposed by Marshalek, Lohman, & Snow (1983).

While the Marshalek et al. (1983) model is useful for capturing the relations among a great many ability tests, a shortcoming of the structure is that it does not take the speed of processing into account. Ackerman (1988) has proposed a modification of this structure that adds speed of processing as a third dimension, orthogonal to the dimensions described by content and complexity. This model is illustrated in Figure 2.

As shown in the Ackerman model, adding the speed dimension has the advantage of delineating unspeeded types of information processing (such as are required in power tests), from speeded information processing that is required in prototypical perceptual speed and psychomotor ability tests. While the Marshalek et al. model would relegate both highly speeded tests and tests that involve specific content (e.g. narrow tests of spatial or verbal knowledge) to the periphery of the circle, there is now an explicit separation of these two classes of ability measures. As will be discussed below, the Ackerman modification enables linkage between ability constructs and phases of skill acquisition. It is important to keep in mind that this structure of abilities not only consistent with the Marshalek theory but is consistent with the broad hierarchical theories of intelligence discussed previously.

IV. Theoretical and Empirical Background

A. Cognitive Ability Determinants of Skill Acquisition

Using the basic outline of skill acquisition reviewed above in conjunction with the ability structure delineated by Ackerman, it is possible to build a representation of the determinants of individual differences in performance during skill acquisition. The structure adopted here is based on the theory proposed by Ackerman (1988) that maps classes of cognitive/ intellectual abilities to the phases of skill acquisition. The theory states that a direct mapping of ability classes and phases of skill acquisition can be established. From this mapping, it then becomes possible to predict the association between individual differences in abilities and individual differences in task performance during skill acquisition. The principles are as follows:

Phase 1. Given the fact that the first phase of skill acquisition has substantial demands on the cognitive system (mostly in the domain of attention, memory, and reasoning; but also in the area of broad content knowledge), this phase of skill acquisition closely represents the types of information processing that are sampled in tests of high complexity and minimal speed demands. Thus, general and broad content abilities (namely, spatial, verbal and numerical) illustrated at the top of the ability cylinder (shown in Figure 2) are implicated in predicting individual differences at the declarative knowledge phase of skill acquisition.

Phase 2. Once the learner has developed appropriate strategies for successful task accomplishment, and moves to the knowledge compilation (or associative) phase of skill acquisition, the demands on the general and broad content abilities are reduced in a fashion analogous to the reduction in demands on the attentional and declarative knowledge systems.

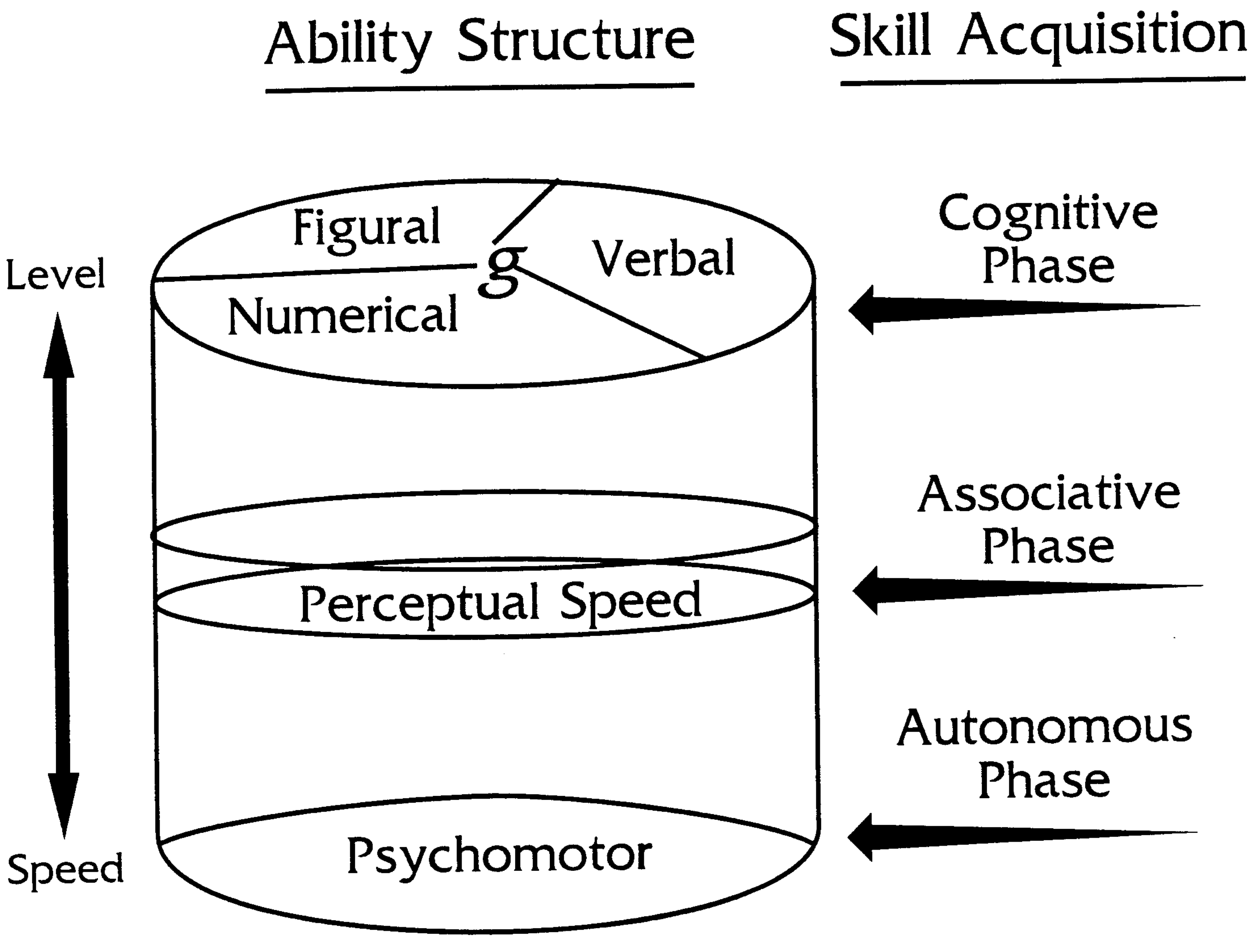


Figure 2. A modified radex-based model of cognitive abilities. Complexity is represented as in the Marshalek, et al. model. However, the dimension of level/speed is added to represent perceptual speed and psychomotor abilities. (From Ackerman, 1988).

Superior performance at this phase of practice is indicated when learners can efficiently develop streamlined productions for task accomplishment, those that improve speed and accuracy. Similarly, as one moves forward on the speed dimension of the proposed ability structure, from general and broad content abilities to perceptual speed abilities, marker measures require fewer demands on attention and memory and greater demands on the refinement of relatively simple productions (as in a digit-symbol test, or a clerical checking test -- both such tests are prototypical exemplars of Perceptual Speed ability). In fact, perceptual speed ability can be thought of as "the capacity to automatize, by means of practice" (Werdelin and Stjernberg, 1969, p. 192). Thus, individual differences in performance at the second phase of skill acquisition will be well-predicted by individual differences in the perceptual speed ability.

Phase 3. When the final phase of skill acquisition, procedural knowledge, is reached, the learner can often effectively meet the task demands with little or no attentional effort (e.g. see Schneider and Fisk, 1982). Performance limitations at this phase, especially when the task is critically dependent on motor operations, are primarily determined by differences in the speed of encoding and responding, with very little involvement of declarative knowledge. The phenomenological essence of such levels of skill acquisition is that task performance is "automatic." The highly deterministic nature of the task has a number of striking similarities to ability measures at the extreme level of speededness, that is, the family of psychomotor tests (such as simple reaction time, or tapping speed). In both task and test scenarios, the individual knows exactly what response (or response sequence) needs to be made once the stimulus is detected. As such, the theory specifies that individual differences in psychomotor abilities will well-predict individual differences in performance at this last phase of skill acquisition.

The overall implication of the theory is that the determinants of individual differences in performance are *dynamic*, that is, different classes of abilities are most highly correlated with performance at each phase of skill acquisition. Figure 3 illustrates the anticipated pattern of ability - performance correlations across a sequence of task practice, leading to proceduralized knowledge. As indicated in the figure, individual differences in performance at Phase 1 are best predicted by general and broad content abilities; at Phase 2, the best predictions come from measures of perceptual speed ability; and at Phase 3, the most salient correlations are found for psychomotor abilities.

B. *Self-Regulatory Processes/Motivation and an Integrated Model*

Motivation refers to: (a) the *direction* of attentional effort; (b) the *proportion* of one's total attentional effort directed to the task (intensity), and (c) the extent to which attentional effort toward the task is *maintained over time* (Campbell & Pritchard, 1976; Humphreys & Revelle, 1984; Kanfer, F. 1987; Kanfer, R., 1991; Kleinbeck, 1987; Naylor, Pritchard, & Ilgen, 1980). From a resource perspective, motivation is defined as the set of processes underlying the volitional allocation of limited cognitive resources, or attentional effort. This definition of motivation, emphasizes metacognitive processes, or the self-regulation of cognitive resources, and allows us to recast motivational phenomena into an integrated information-processing resource-allocation framework.

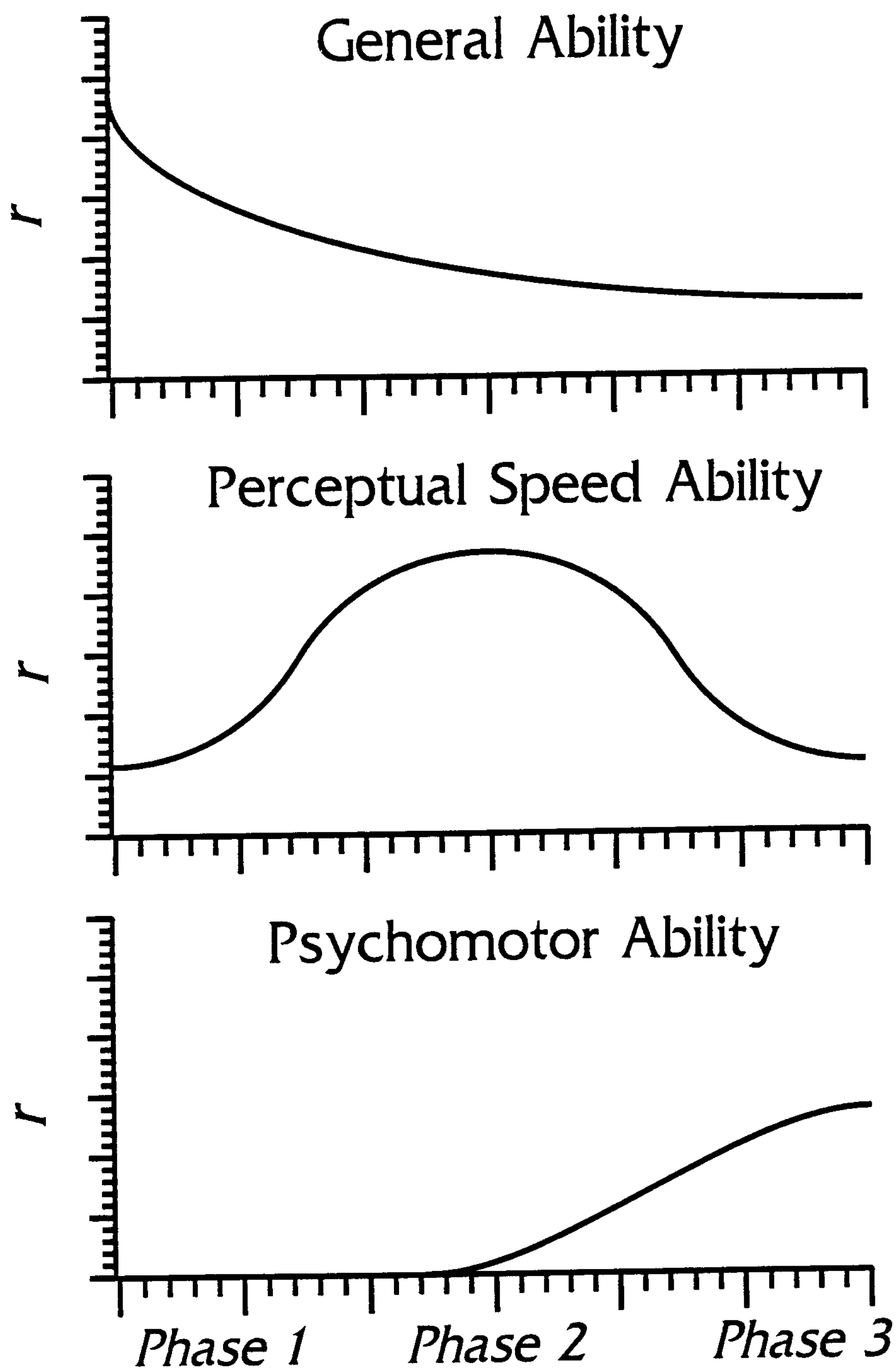


Figure 3. Hypothetical ability/skill relations derived from the framework. The hypothetical task is moderately complex, involves a moderate amount of broad transfer, and provides for consistent information processing. (From Ackerman, 1988).

Kanfer (1987, 1991) suggested that motivation affects learning and task performance through its influence on two distinct resource-allocation processes. Distal resource-allocation processes guide the establishment of an individual's behavioral intentions and choice among goals. Learners develop specific intentions (e.g., to learn a specific skill) and choose behavioral objectives (e.g., to obtain a specific performance score). The resource allocation processes underlying the development of these conscious intentions and goals are termed *distal*, because their effects on performance are indirect and depend on how the intentions and goals are implemented during learning.

Research on the effects of distal motivational processes has emphasized prediction of the effects of motivation on intended effort rather than the effects on skill acquisition or task performance per se (Locke, 1968; Mitchell, 1982; Vroom, 1964). As a result, information-processing demands of the task, individual differences in abilities, and other self-regulatory processes that may mediate the intention - performance relation have previously been neglected in these formulations.

When a learner's goals are readily attainable or anticipated to be easily achieved, goal choices made through the distal motivational system are typically realized quickly and without difficulty. However, when goals involve complex or novel learning, goal attainment typically requires additional volitional, or metacognitive activity to guide resource allocations during skill acquisition. The second source of motivational effects on skill acquisition stems from resource-allocation processes that occur *during* learning. These motivational processes are termed *proximal* since their operation has direct consequences for learning and goal attainment. Such motivational processes include metacognitive knowledge and the self-regulatory activities by which the learner directs, energizes, and sustains attentional effort for the purpose of learning. Whereas distal allocation processes influence the learner's decision to exert effort, proximal motivational processes operate in the context of competing demands for attentional effort imposed by the task and limitations of resource availability attributable to individual differences in cognitive ability. Metacognitive knowledge and self-regulatory activities can have a direct effect on rate and asymptotic level of skill acquisition. Proximal motivational processes that promote greater allocations of attentional effort to task components facilitate skill acquisition.

A cognitive resource allocation approach to motivation permits a distinction between motivational processes affecting an individual's decision to exert effort (distal) and the volitional processes by which intentions are translated into performance (proximal). Distal motivational processes typically occur prior to, or following task engagement. In contrast, metacognitive and self-regulatory processes *operate within the context of attentional demands imposed by the task and constraints on total resource availability*. Proximal resource allocation processes occur during learning and thus may affect performance through their effects on skill acquisition and effort.

An important implication of the distinction between distal and proximal motivation processes pertains to understanding how metacognitive processes and self-regulation affect skill acquisition. A number of studies demonstrate the beneficial consequences of self-regulation -- increased effort (i.e., increases in the total proportion of an individual's resource capacity devoted to the activity) and/or focusing a greater proportion of allotted resources to on-task activities in well-learned tasks (see Bandura, 1986; Kanfer, 1977). However, as Kluwe and Friedrichsen (1985) suggest, self-regulation itself requires cognitive/attentional resources. To obtain the beneficial consequences of self-regulation there must be cognitive resources available for engaging in self-regulatory activity. When self-regulatory activities demand resources that can *only* be provided through a reduction in resources demanded by the task, operation of self-regulatory activities may exert a cognitive cost that impairs learning and performance. This situation (in which tasks demand almost full use of available cognitive resources) frequently occurs when persons first encounter a difficult or complex task.

By integrating self-regulation demands and cognitive resources, it is possible to consider self-regulation in the context of skill acquisition. This perspective suggests that the engagement of self-regulatory activities when the task is resource dependent (e.g., during the declarative phase of skill acquisition) will deprive the task of needed resources. Unless the benefits of self-regulation are stronger than the costs of resource diversion, performance (and subsequent learning) will suffer. Following the development of a declarative representation of the task (i.e., in the knowledge compilation or procedural phases of skill acquisition) however, the engagement of self-regulatory activities are expected to enhance performance. This benefit appears to be due to the availability of cognitive resources (for self-regulatory activity) and the consequences of self-regulation that serve to increase resource allocation to on-task activity.

Toward a unified model of skill acquisition

By mapping abilities and motivation to performance-resource functions (Norman & Bobrow, 1975), a learner's performance may be represented as a joint function of the learner's relative attentional capacity (i.e., cognitive ability), task demands (i.e., phase of skill acquisition), and the proportion of the learner's total capacity actually devoted to the task (i.e., motivation). Figure 4 illustrates the integrated framework underlying the planned research. As shown, the model represents a modification and elaboration of Kahneman's (1973) model of attentional capacity (Kahneman, 1973; Figure 1-2, p. 10). Attentional capacity is viewed as an interindividual differences attribute. Attentional resources are allocated across different activities; feedback loops are posited for adjustment of allocations, proportion of total capacity allocated, and for external influences at both the level of allocation of capacity and allocation policy. In contrast to Kahneman, however, our representation explicitly distinguishes among three types of possible activities: (a) off-task activities, (b) on-task activities, and (c) metacognitive/self-regulatory activities. A basic assumption underlying our model is that changes in the amount of capacity utilized and policies for allocation of attention are accomplished through motivational processes.

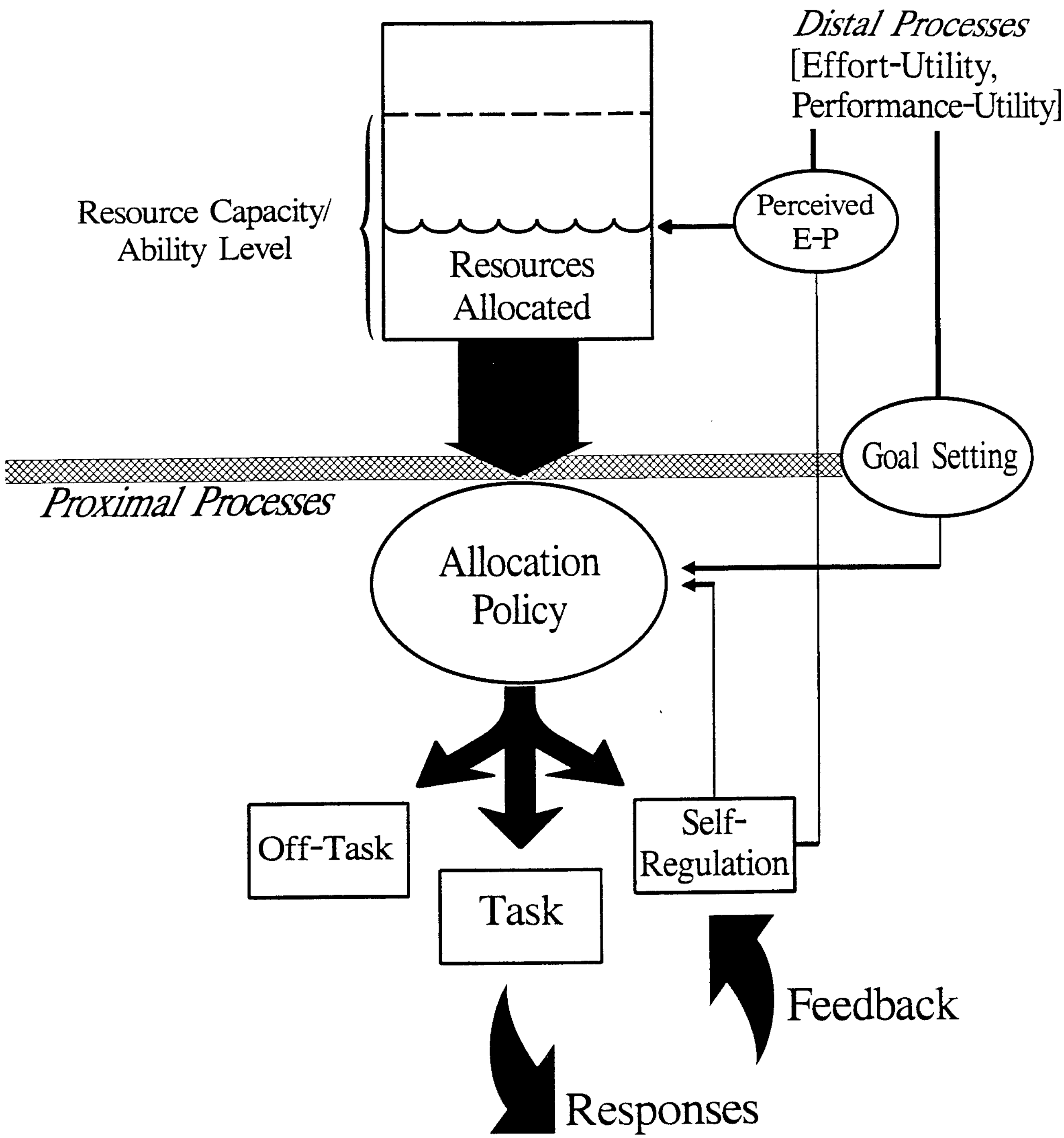


Figure 4. The Kanfer-Ackerman (1989) model of ability/motivation interactions for attentional effort. The model is derived from a model of attention proposed by Kahneman (1973).

The resource framework also provides a means of conceptualizing dynamic changes in performance as a function of abilities, motivation, and task characteristics. Over time, distal and proximal motivational processes influence the re-allocation and/or mobilization of additional portions of capacity to various activities. When task resource demands are high, the individual may allocate more available attention to the task, or adjust the proportion of capacity engaged. Conversely, when task demands are reduced, the individual may attend to off-task activities and/or may reduce the proportion of capacity engaged.

In this model, self-regulation is an essential mechanism for bringing about changes in allocation policy toward a task or total proportion of resource capacity actually engaged. Without activation of self-regulatory processes, for example, performance feedback would have little influence on learning and the individual would be expected to continue to devote the same amount of resources originally committed to a task from distal decision processes.

Individual differences in ability level/resource capacity will determine the total amount of resources that can be devoted to any set of activities. Consistent with Figure 4, low-ability learners must devote a greater portion of their capacity than high ability learners in order to achieve similar levels of task performance.

However, Figure 4 is a static representation of ability and motivational processes. Changes in the performance-resource function that occur during skill acquisition are not directly represented in the figure but are addressed in the integrated framework. As noted previously, a potential drawback to activation of self-regulatory processes is that such processes may draw away resources needed to develop a declarative representation of the task. In this initial, resource dependent phase of learning, self-regulatory activity itself represents a "cognitive cost" and the products of self-regulation cannot be readily realized.

Changes in the true performance-resource function are posited to be detected by self-regulatory processes (i.e., self-monitoring of performance feedback and effort allocations). As the task becomes less resource-dependent (over practice), learners discover that fewer resources need to be devoted to the task to maintain performance. As such, resources might be diverted to other activities, including additional self-regulatory processing. Alternatively, an advanced learner may choose to reduce the proportion of attentional capacity currently engaged. To the degree that self-regulatory activities in tasks approaching resource-insensitivity not only enables motivational processing without costs but further increases, directs, and sustains on-task attentional effort, external influences that prompt self-regulation as task resource demands decrease may enhance performance. Taking into consideration individual differences in ability, the detrimental and beneficial effects of self-regulatory activity in the declarative and later phases of skill acquisition, respectively, are expected to be greater for low-ability learners than high-ability learners.

V. Study 1 - Target/Threat Identification Task - Laboratory

Introduction and Overview

This study was designed to accomplish the following: (1) Prospective assessment of cognitive ability and self-regulatory skill determinants of stress-reactivity; (2) Concurrent measurement of performance under stress-induced conditions; and (3) Retrospective measurement of self-regulatory processes during information processing and decision-making (e.g., ongoing self-monitoring, self-evaluation, and self-reactions).

For the first component of the study, tests of cognitive and perceptual abilities, personality, self-concept, self-ratings of ability, and self-regulation were administered. For the second component of the study, the Task/Threat Identification Task was used for concurrent measurement of performance under low and high stress conditions. Stress was manipulated in terms of time-pressure and information processing demands. The basic demands of the task (rapid categorization, spatial scanning, information accumulation under time-pressure, ability to take advantage of consistent information) greatly overlap those in the CIC situation, and this instantiation of the task is feasible for testing in the limited (2-3 hour/session) time frame. The third component of the study included performance on a short training session with a highly complex information processing task -- a simulation of air traffic control (Terminal Radar Approach Control, TRACON). The fourth component of the study provided assessment of task-relevant self-regulatory processes during information processing and decision-making.

Overview of Measures

This experiment contains three major categories of measures: Distal individual differences measures, Proximal individual differences measures and Performance measures (Target/Threat Task, TRACON). These measures are described briefly below, and are presented in greater detail in the Method section.

1. *Distal Measures.* Numerous distal measures were administered, in order to converge on the multiple determinants of performance in the Target/Threat and TRACON simulation tasks. The measures were designed to assess individual differences cognitive/intellectual abilities, personality, motivational skills, and self-ratings of ability.

2. *Proximal Measures.* A questionnaire (Interim Questionnaire) was administered prior to each Target/Threat task session and the TRACON task session. This questionnaire assessed aspects of task-specific Self-Efficacy.

3. *Task Performance Measures.*

Target/Threat. Numerous measures were collected during the Target/Threat task (including score, A' [a measure of signal-detection sensitivity], average reaction time [RT]). During dual-task conditions, an additional secondary task memory-accuracy measure was collected.

TRACON. As in previous studies, the major measures used to assess performance in TRACON relate to the number of planes successfully handled (carried to final disposition) during each simulation. Given only a limited amount of task practice was provided, only overall performance (across six task trials) was examined.

Subjects

Subjects were recruited by signup sheets and flyers distributed around the University of Minnesota. Subjects were restricted to: (a) age between 18 and 30 years, (b) normal or corrected-to-normal vision, hearing, and motor coordination; and (c) no prior experience with ATC tasks. One hundred and eight subjects participated in the experiment. Of these subjects, 63 were males, and the age level for the whole sample was: $\underline{M}$ = 20.5 years, $\underline{sd}$ = 2.92 years. Subjects were paid $75 each for the 12 hr of participation. Payment was not contingent on task or test performance.

Apparatus

The computerized self-report measures, Target/Threat task and TRACON trials were administered on Compaq Deskpro computers with Intel 80386/20MHz SX processors and NEC 4D monitors. Information was displayed in text mode (for the questionnaires), or in color VGA (680 H x 480 V) resolution graphics (for the Target/Threat and TRACON trials). Input was accomplished by keyboard responses, or with a Logitech 3-Button mouse (Target/Threat task) or MicroTech 3-button trackball (TRACON). Audio information for the Target/Threat Dual-Task was presented with a cassette tape and a public address system. Audio information (for TRACON) was presented binaurally over headphones to each subject (using a SoundBlaster interface). Each computer was housed on a separate table, shielded visually and aurally by large acoustic panels that enclosed each computer. For ability testing, subjects were run in groups up to 30 at a time. For computer Target/Threat and TRACON task practice, subjects were run in groups up to 8 at a time.

Ability Tests

One battery of tests was administered using a paper and pencil format: The battery contains 13 ability tests and an 18-item measure of motivational skills, in the following order: (1) Verbal Analogy; (2) Digit-Symbol; (3) Number Comparison; (4) Paper Folding; (5) Coding, (6) Number Series, (7) Flight Checking, (8) Spatial Orientation, (9) Name Comparison, (10) Spatial Analogy, (11) Name Symbols, (12) Verbal Test of Spatial Ability, and (13) Dial Reading. (The Motivational Skills measure follows the ability tests in the administration order). Tests #1, 4, 7, 8, 10, 12, 13 are from the Aptitude Assessment Battery (AAB). Descriptions of the AAB tests can be found in Ackerman & Kanfer (1993). The remaining tests were either commercial tests (e.g., Number Series) or were locally developed tests. Additional tests were presented on the computer: Plane Memory (from the AAB) and 9-choice RT, 4-choice RT, and 2-choice RT tests.

Personality Measures

In order to allow investigation of the relations between the criterion performance and the broad personality domain, the NEO-PIR (Costa and McCrae, 1992) was administered. The NEO-PIR is a 240 item inventory designed to assess standing on five broad factors of personality. The five broad personality factors assessed by the NEO-PIR are: Extroversion, Neuroticism, Openness, Agreeableness, and Conscientiousness. Additional personality measures included Impulsivity, Achievement Motivation (WOFO, Helmreich & Spence, 1978), and Trait Anxiety (Spielberger, 1982), Action Control (Kuhl, 1985), and Control (MPQ, Tellegen, 1982).

Self-Ratings of Ability

For self-ratings of ability, a 21-item questionnaire was used. The items represent three broad factors of abilities: self-regulation ability, cognitive ability, and clerical ability. Subjects were instructed to respond with a self-evaluation *relative to other persons their age*.

Motivational Skills

The Motivational Skills measure is an 18-item questionnaire assessing aspects of self-confidence for learning, studying, and performing in test situations (for a description of the measure see, Ackerman and Kanfer, 1993). The Motivational Skills questionnaire has been shown to provide incremental prediction of complex task performance after abilities are partialled out.

Self-Efficacy Measures

A series of self-report questionnaires were administered before the Target/Threat trials and TRACON simulation trials on Sessions 2 - 4. Trainees were asked to indicate their self-efficacy for performance on the Target/Threat task and TRACON task performance (overall performance). Questions were administered in an ordered fashion. Specifically, SE_1 was administered prior to the first Target/Threat Task (Single Task Conditions), SE_2 was administered prior to the Target/Threat Dual-Task conditions, and SE_3 was administered prior to the TRACON task. Questions for both tasks referred to performance on the current session, for example: "I can handle *xx* percent of the planes entering the airspace" (where *xx* refers to 20, 40, 60, 80, and 100).

Target/Threat Identification Task. The task is described in **Appendix A**.

Terminal Radar Approach Control (TRACON). The task is described in Appendix B.

Procedure

The experiment was completed in 4 sessions: Session 1 contained only ability testing and self-report measures with paper and pencil format; Session 2 contained computerized ability tests, computerized self-report measures, instructions and the first two sessions of Target/Threat task; Session 3 was a dual-task Target/Threat task session, and also contained a TRACON training video; and Session 4 was devoted entirely to TRACON practice, spread out over a one week period.

For the Target/Threat task in Session 2, trainees first completed an interactive computerized instruction program on the target/threat task. The program demonstrated the use of the mouse, showed the layout of the task and feedback screens, and presented three practice trials of the target/threat task (with a smaller number of targets and ships). Trainees completed the first interim questionnaire, and then completed 30 trials of the CM version of the target/threat task, completed the memory test, and completed 30 trials of VM version of the target/threat task. At the end of the VM trials, trainees completed the first task perceptions questionnaire.

For the Target/Threat task in Session 3, trainees were given a 1 minute demonstration of the audio task (the secondary task), the second interim questionnaire, and then performed 30 trials of the CM target threat Dual-Task, completed the target/threat memory test, the Identifier Recall Task (from the secondary task), and then completed the VM target/threat dual-task, and completed the Identifier Recall task (from the secondary task). Trainees were then administered the second task perceptions questionnaire. After a break, trainees watched the video instruction for the TRACON task.

For the TRACON task in Session 3, trainees first completed the interim questionnaire, and then performed 6, 30-minute TRACON simulations. After the TRACON simulations, trainees completed the final task perceptions questionnaire.

Results

Target/Threat Performance. Although numerous task performance measures were collected and computed, all of the general performance measures showed convergent validity (i.e., were highly correlated). As such, brevity is obtained by limiting discussion to a single performance measure, which for current purposes was selected to be RT (mean response time from hooking targets to entering an identification response). Performance on the four conditions (across trials) was: (1) CM Single M = 1,730, sd = 429, (2) VM Single M = 1,985, sd = 292, (3) CM Dual M = 1,332, sd = 388, and (4) VM Dual M = 1,750, sd = 268. As in other lookup-type tasks, interindividual differences (*sd*s) were higher in CM conditions, and mean RTs were higher in VM conditions (e.g., see Ackerman & Woltz, in press). In addition, the reduction of mean RTs from single-to-dual conditions indicates a general effect of practice on the Target/Threat task.

Ancillary measures. Memory performance on the two CM tasks was generally quite good (CM Single M = 8.85, *sd* = 1.49; CM Dual *M* = 9.34, *sd* = 1.06), indicating a mean of 88% accuracy in the CM single task, and 93% accuracy in the CM dual task. For the two dual-task conditions, there was a general improvement across the two sessions (CM Dual *M* = 2.5, *sd* = 2.79, VM Dual *M* = 4.8, *sd* = 3.47).

TRACON Performance. Only one "session" of TRACON practice was completed by the trainees. Overall performance (mean number of planes handled across simulations) was consistent with previous studies. Across six simulations, mean performance was *M* = 7.70, *sd* = 3.85.

Ability - Performance Correlations

Target/Threat Task. Correlations between ability composites (Reasoning, Spatial, Associative Memory, Perceptual Speed, and Reaction Time (RT)) were computed for each task condition, as were correlations for the Dial Reading Test, and are shown in Figure 5 (for the CM conditions) and Figure 6 for the VM conditions. In addition, ability - performance correlations were computed for the memory tests in the two CM conditions, and are shown in Figure 7. Consistent with expectations, the two CM conditions showed highest correlations with the Associative Memory composite, and with the Reaction Time composite, reflecting the two main components of the task: Associative Memory ability is associated with the facility of learning the target-identification associations, which in turns speeds up overall performance on the task; Reaction Time is associated somewhat with overall facility in using the mouse and mostly with making quick and accurate keyboard/mouse key responses. The pattern for the two VM conditions was different from that of the CM conditions. In the VM condition, Perceptual Speed ability was also substantially associated with task performance, as expected, because perceptual speed is required for rapid scanning of the lookup table. These results are consistent with substitution tasks that are isomorphic in process, but only use static stimulus displays (Ackerman & Woltz, in press).

For the memory tests, Reasoning and Spatial abilities are important determinants of performance (which is concordant with the fact that these abilities typically are identified with a general ability factor), but the highest correlations are with the Associative Memory factor, in line with the processes that are required for this task component.

Generally, the ability-performance correlations for the Target/Threat task confirm the task characteristics that were intended by the designers. However, the dual-task versions of the Target/Threat task did not show consistent elevations in ability - performance relations that were expected. The apparent explanation of these results comes from an examination of the respective mean performance scores across the four task conditions. That is, general practice effects were shown across the sessions, such that overall performance improved, even as the task information processing requirements increased from single-to-dual task conditions. As such, the improved levels of performance appear to reflect a lack of increase in task difficulty from single-to-dual conditions, and thus no general increase in ability-performance correlations.

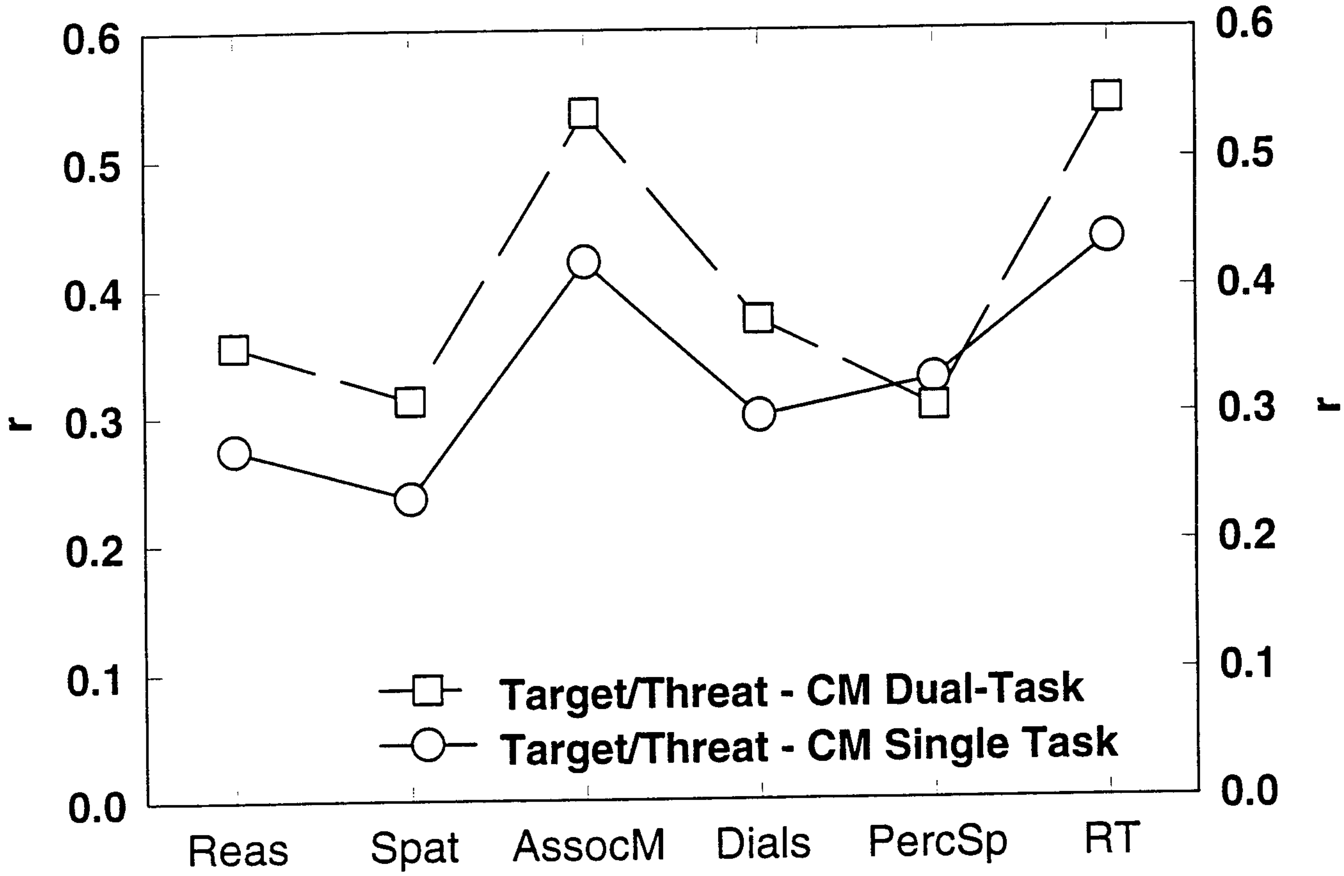


Figure 5. Correlations between abilities and performance on CM Single-Task and Dual-Task versions of the Target/Threat Identification Task. (NOTE: Reas = Reasoning, Spat = Spatial, AssocM = Associative Memory, Dials = Dial Reading Test, PercSp = Perceptual Speed, and RT = Reaction Time.)

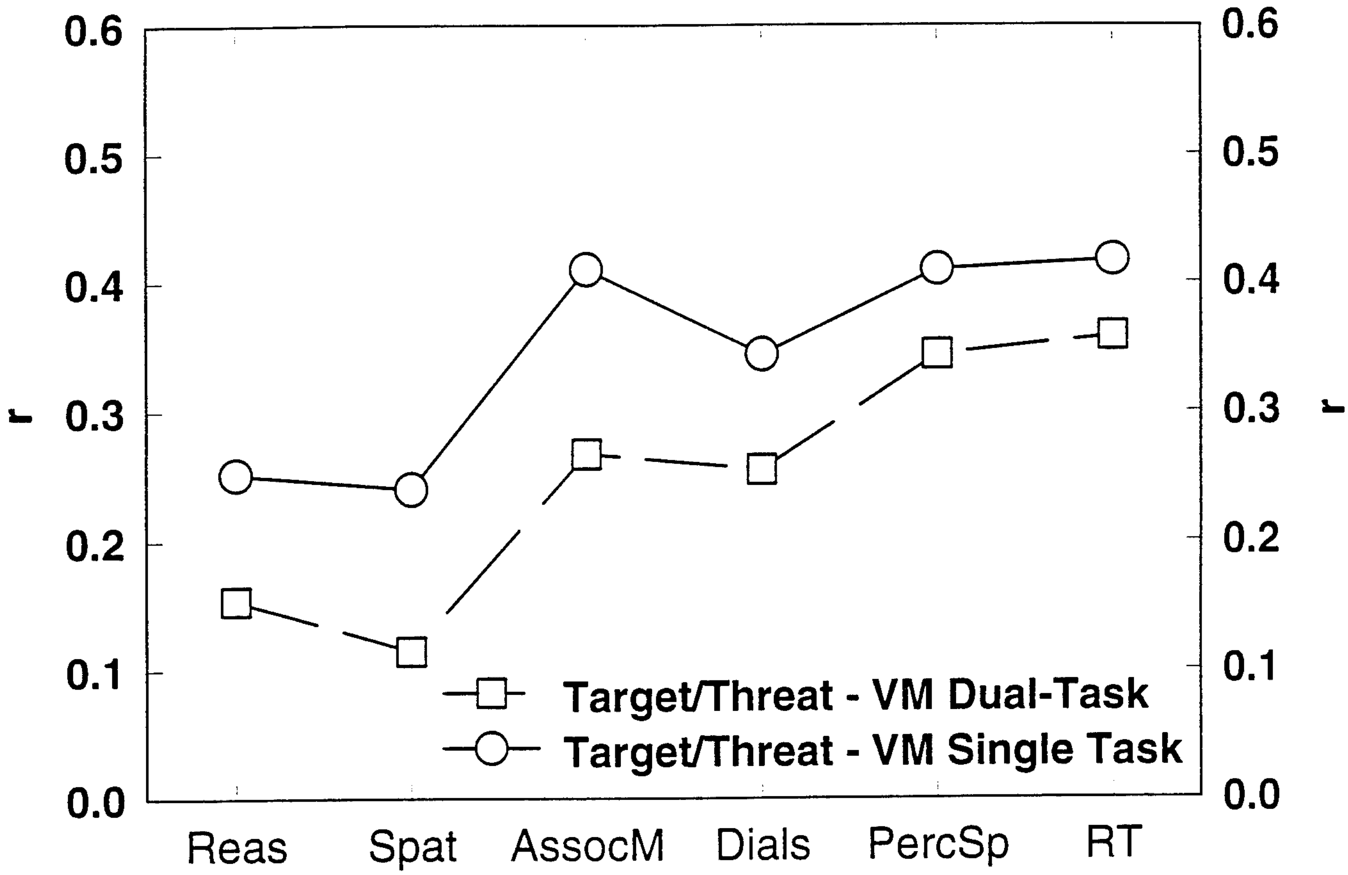


Figure 6. Correlations between abilities and performance on VM Single-Task and Dual-Task versions of the Target/Threat Identification Task. (NOTE: Reas = Reasoning, Spat = Spatial, AssocM = Associative Memory, Dials = Dial Reading Test, PercSp = Perceptual Speed, and RT = Reaction Time.)

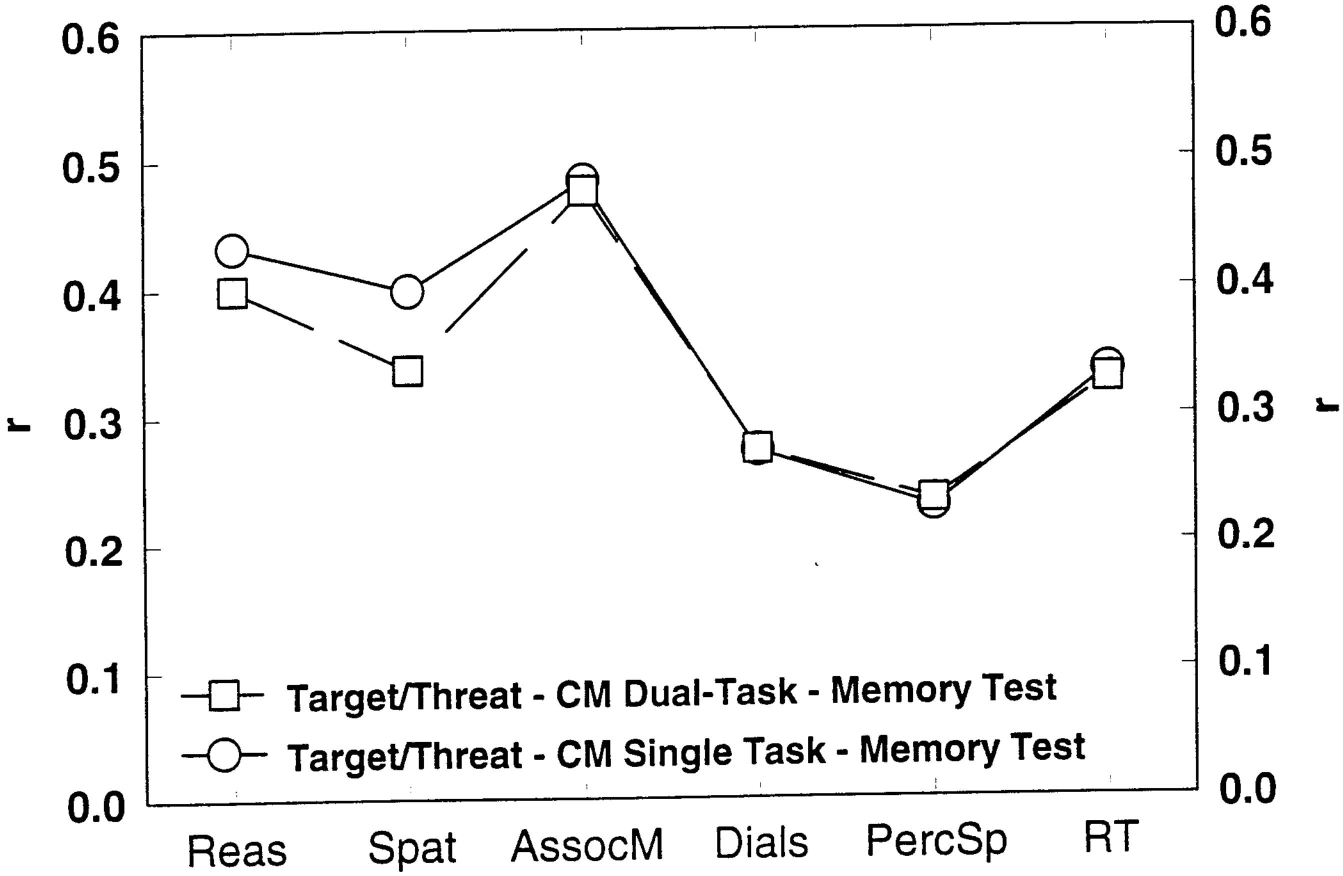


Figure 7. Correlations between abilities and performance on CM Single-Task and Dual-Task versions of the Target/Threat Identification Task -- Memory Tests. (NOTE: Reas = Reasoning, Spat = Spatial, AssocM = Associative Memory, Dials = Dial Reading Test, PercSp = Perceptual Speed, and RT = Reaction Time.)

TRACON Performance. Correlations between ability composites and TRACON performance are shown in Figure 8. Consistent with previous TRACON studies, the highest correlations were found for Spatial and Reasoning ability composites.

Non-Ability - Performance Associations

Personality Measures. The broad measures of personality showed no consistent significant correlations with task performance (both RT and Memory) in the Target/Threat task conditions. Only the Neuroticism significantly correlated with TRACON performance ($r = -.26$, $p < .01$).

Motivational Skills. As in previous studies, the Motivation Skills measure showed a significant correlation with TRACON performance ($r = .26$, $p < .01$). In addition, this measure also showed small, but significant correlations with the Memory Test performance for both CM tasks ($r = .22$, $p < .05$, for both CM Single and CM Dual-Task conditions).

Self-Ratings of Ability. The self-ratings of ability measure was factor analyzed to reveal three general factors. Factor 1 was identified as Self-Regulation -- it included items of planning ability, self-control skills, personal organizational skills, and self-discipline skills. Factor 2 was identified as a Cognitive Ability -- it included items of math, spatial, mechanical and problem-solving ability, and perceptual skills. Factor 3 was identified as Clerical Ability -- it included items of office skills, dexterity skills and clerical skills. Composite scores were computed to reflect these three general factors of self-ratings of ability, and were then correlated with performance variables. The Self-Regulation composite and the Clerical Ability showed no consistent significant correlations with any of the performance measures. In contrast, the Cognitive Ability self-rating composite had significant correlations with TRACON performance ($r = .30$, $p < .01$), and with the Memory Test scores for the CM Target/Threat conditions ($r = .21$ and $.23$, $p < .05$ with CM Single and CM Dual Task conditions respectively).

Self-Efficacy Measures. Task specific self-efficacy measures were administered prior to each of the three task sessions (but after instructions had been presented for each task). Specifically, SE_1 was administered prior to the first Target/Threat Task (Single Task Conditions), SE_2 was administered prior to the Target/Threat Dual-Task conditions, and SE_3 was administered prior to the TRACON task.

SE_1 demonstrated significant correlations for the CM Single Task condition -- RT ($r = .21$, $p < .05$) and Memory Test ($r = .20$. $p < .05$), a marginal correlation with VM Single Task RT ($r = .17$, $p = .09$), but did show a significant correlation with later TRACON performance ($r = .32$, $p < .01$).

SE_2 demonstrated significant correlations for the CM and VM Dual-Task conditions -- RT ($r = .33$, $.23$ respectively, $p < .05$) and Memory Test ($r = .23$, $p < .05$). In addition SE_2 also showed a significant correlation with later TRACON performance ($r = .34$, $p < .01$).

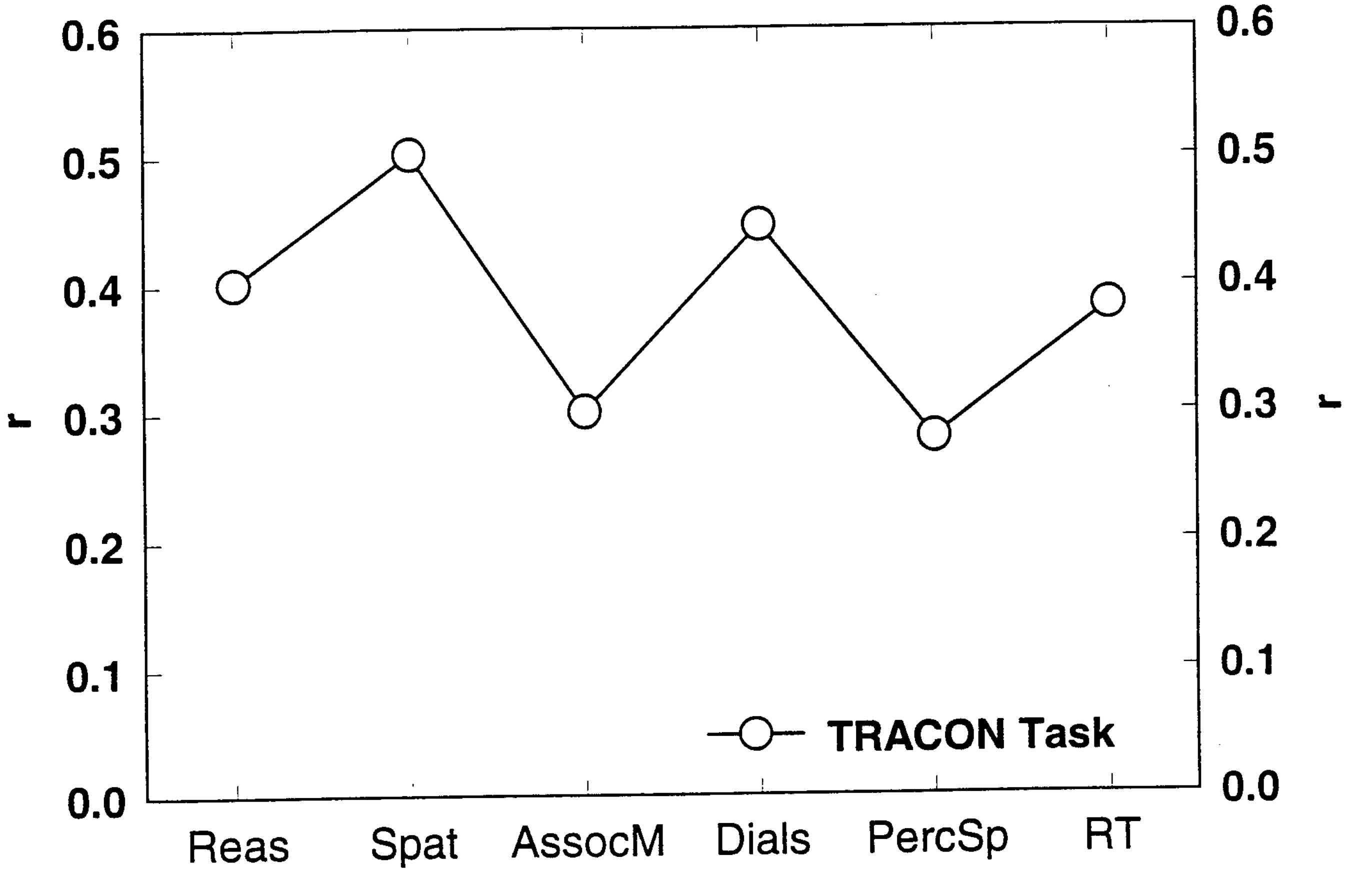


Figure 8. Correlations between abilities and performance on TRACON. (NOTE: Reas = Reasoning, Spat = Spatial, AssocM = Associative Memory, Dials = Dial Reading Test, PercSp = Perceptual Speed, and RT = Reaction Time.)

SE_3 demonstrated only a significant correlation with TRACON performance (r = .40 $p < .01$).

Gender Effects

Significant gender effects were not consistently demonstrated for the Target/Threat task conditions, either for RT or for Memory Test performance. However, consistent with previous studies, a significant difference between males and females was found for TRACON performance (M_{males} = 8.67, sd_{males} = 4.06; M_{females} = 6.35, sd_{females} = 3.10, $t(106)$ = 3.23, $p < .01$).

Multiple Correlations (Overall Prediction)

A series of hierarchical multiple correlations were computed for each of the Target/Threat Conditions and for overall TRACON performance. The results of these analyses is presented in Table 1. In any hierarchical regression, the order of entry for predictor measures is admittedly somewhat arbitrary. However, we decided to enter the cognitive/intellectual ability measures first, given that cognitive ability is quite stable in adults, and that cognitive ability measures are often readily accessible in the field. After ability was entered, additional measures were entered in each step: Motivational Skills, Self-Estimates of Ability, Personality, and Self-Efficacy. Finally, we added gender as the final variable, to determine whether any observed significant gender differences in performance could be first accounted for by gender differences in the predictor variables.

The results of the multiple correlations across the tasks and conditions are relatively clear. First, with a relatively small number of predictor variables, we were able to account for an average of about 30% of the variance in performance on the Target/Threat task conditions, and about 40% of the variance in performance on the more complex TRACON task. Second, with no exceptions, individual differences in cognitive/intellectual abilities accounted for the largest portion of the explained variance in task performance -- regardless of task consistency, or the stress-inductions of dual-task or complex task performance. In several task conditions, self-estimates of ability accounted for additional variance (though making a significant incremental validity in only the most complex task, that is, TRACON performance). Personality measures accounted for only a very small (and non-significant amount) of variance in performance, *after abilities were entered first into the prediction equations*. Finally, as we are beginning to see across a number of tasks (Kanfer, 1994), task-specific self-efficacy, measured prior to task engagement, adds little or no significant predictive validity for task performance, again *after abilities are first entered into the prediction equation*. Similarly, gender did not add significant prediction to the equation, suggesting that for these task conditions, any observed gender differences could be accounted for by gender differences on the other predictor measures.

Table 1. Hierarchical Multiple Regression/Correlation Results for Study 1. Variance accounted for.

Predictor	Criterion				
	CM Single	VM Single	CM Dual	VM Dual	TRACON
Ability	25.0%**	23.8%**	32.3%**	19.0%**	28.6%**
Motivational Skills	0.0	0.0	2.3	0.3	2.1
Self-estimates of ability	2.6	5.6	0.6	4.0	5.6*
Personality	2.1	0.5	0.2	0.4	2.7
Self-Efficacy (pre-task)	0.3	0.0	0.4	0.8	2.8*
Gender	2.5	0.0	1.3	0.0	0.0
Total Variance Accounted For	33.0	29.9	37.3	24.0	41.9

Note. * $p < .05$; ** $p < .01$.

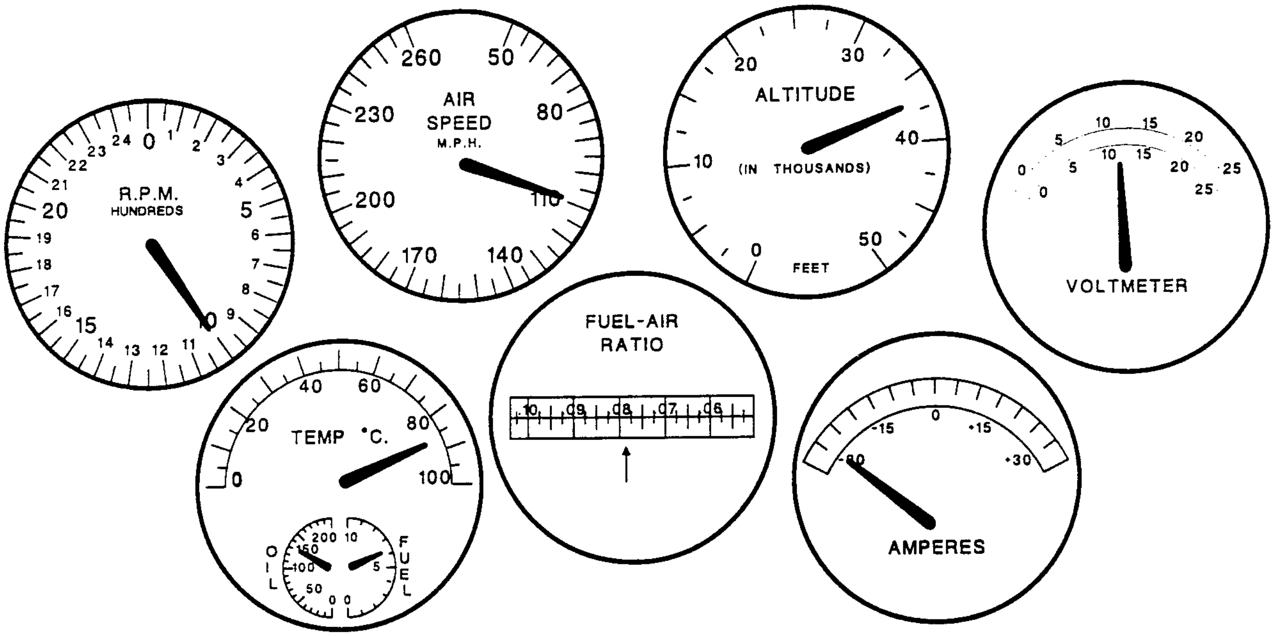
Dial Reading Test

The reason for separately presenting the Dial Reading Test correlations with performance measures (see Figures 5 to 8) was two-fold. First, the Dial Reading Test has been a long history of consistently showing high validity for job tasks that involve a variety of stressors (e.g., Military pilot and navigator jobs, see Guilford & Lacey, 1947; air traffic controllers, see Cobb & Mathews, 1972, and Ackerman & Kanfer, 1993). Second, the test also revealed an interesting pattern of significant correlations with non-ability predictor measures. In the current study, the Dial Reading Test significantly correlated with all measures of performance (CM Single and Dual, VM Single and Dual, Memory Tests, secondary task performance, and TRACON performance). Moreover, the Dial Reading Test significantly correlated with personality measures of Openness (r = -.21) and Agreeableness (r = -.21), and with the self-rating of cognitive ability (r = .25). Finally, the Dial Reading Test significantly correlated with all three measures of self-efficacy (SE_1 = .35, SE_2 = .31, and SE_3 = .24). The fact that this measure correlated both with task performance under a variety of stressful conditions, and with personality, self-ratings of ability, and self-efficacy, suggested that closer examination of the Dial Reading Test was in order.

The Dial Reading Test was first developed by the U.S. Army Air Force Aviation Psychology Research Program during WWII (Guilford & Lacey, 1947). Originally, this test was part of a more comprehensive Dial and Table Reading Test (the table reading portion was a speeded lookup and interpolation test). Figure 9 shows sample items from the Dial Reading Test. The test has several dials portrayed on each page, with indicator needles on each dial (on each page the dials are constant, but the indicators are placed in different positions on subsequent pages). Each set of dials has an associated set of multiple choice questions, that require the examinee to quickly read (or interpolate) indicators on the dials. Initial analyses of the whole test indicated that it had significant loadings on Numerical, Spatial and Perceptual Speed ability factors. As the investigators noted, "Of all the tests in the classification battery, Dial and Table Reading has the distinction of being the best single predictor of success for any aircrew specialty" (Guilford & Lacey, 1947, p. 403).

Later investigations of the test (with only the Dial Reading Test) by the Federal Aviation Authority suggested that it was also one of the best predictors of success in air traffic controller success (Cobb & Mathews, 1972). *No* information provided by the Army Air Force or the FAA suggested that the test assessed any other factors than those from the ability domain. The correlations between our version of the test and non-ability predictors of performance, though, suggest that there may be something more that cognitive abilities involved in individual differences in test performance.

A task-analysis perspective on this test suggests that, in addition to the spatial and perceptual speed components, there is a component to this test of something like "resistance to interference" under time-stress conditions. That is, when an examinee is confronted with this test, the examinee must make numerous back-and-forth comparisons between the dials and the answer choices, because the choices randomly require either fine interpolations of the dial readings, or only order-of-magnitude readings. For the examinee, it is impossible to



	A	B	C	D	E
Temperature	87.	92.	8.5	91.	81.5
R.P.M.	7.0	0.0	1.0	10.0	10.2
Oil	130.	150.	152.	148.	6.0

Figure 9. Example of items from the Dial Reading Test (from Ackerman & Kanfer, 1993; Original test by the U.S. Army Air Force Aviation Psychology Research Program, Guilford & Lacey (Eds.), 1947).

know, a priori, how accurately he/she must read the particular dial. If the examinee takes the time reading the options thoroughly first, he/she may have wasted valuable time (when the item asks for an interpolated answer). Similarly, if the examinee interpolates the dial reading first (when the item only asks for an order-of-magnitude answer), time is also wasted. As a result, successful examinees must perform rapid back-and-forth looks at the items and the dials, until they converge on the appropriate answer. Examinees that are easily flustered by the time-stress, or the successive back-and-forth processes, perform poorly on this test. In this sense, it appears that successful performance on the Dial Reading Test has influences from outside the cognitive ability domain, namely the domains of personality, motivation, and self-regulation that are also important for performance under stressful conditions. Such an interpretation is consistent with the pattern of significant non-ability correlations observed in this study, and also helps explain why this test is a better predictor of combat aircrew performance and air traffic controller performance than more standard tests of numerical, spatial, and perceptual speed abilities. As such, we included this test in Study 3, along with another test that has similar characteristics (the Directional Headings Test developed by the FAA).

Discussion

The first major result from this study was the set of findings that both ability and non-ability measures were significantly related to individual differences in performance over a variety of task information-processing stressor conditions. However, this result must be viewed in light of the fact that, in a multiple regression/correlation equation, the non-ability predictor measures added no consistent significant increment to prediction of performance under these conditions. That is, common variance among non-ability predictors and criterion measures was accounted for by individual differences in objective cognitive and perceptual ability measures. The other major result from this study concerns the follow-up analysis of the Dial Reading Test, ostensibly a measure of spatial, numerical and perceptual speed ability. Examination of correlations between performance on the Dial Reading Test and non-ability self-report measures indicated that test performance shares a considerable amount of variance with non-ability measures, including personality, self-regulation, and motivational measures. These results suggest that it may be possible to develop an objective, ability-based battery of tests that account for both cognitive ability and non-ability influences upon individual differences in performance under both stressful and non-stressful conditions of information-processing tasks. Although it was not possible to follow-up this hypothesis in Study 2 (because of testing time constraints on the subject sample), we return to this issue in Study 3.

VI. Study 2 - Target/Threat Identification Task - Field

Overview of Measures

Given the limited time constraints in the field sample, this experiment contains two major categories of measures: Distal individual differences measures and Target/Threat performance measures (Threat Task). These measures are described briefly below, and are presented in greater detail in the Method section.

1. *Distal Measures.* A limited number of distal measures was administered, in order to converge on the determinants of performance in the Target/Threat simulation tasks. The measures were designed to assess individual differences cognitive/intellectual abilities, personality, motivational skills, and self-ratings of ability.

2. *Target/Threat Task Performance Measures.*
Measures collected in the Target/Threat task were identical to those in the CM condition of Study 1.

Subjects

Trainees were obtained from a population of U. S. Navy enlisted personnel being studied by Drs. Ross Vickers and Linda Hervig, at the U. S. Navy Health Research Center, in San Diego, CA. Trainees were in the midst of completing their basic preliminary training during the test period. Some trainees failed to complete the study in the time allotted, leaving a total of 227 trainees with complete data. Analyses presented in this report are computed only on trainees with complete data.

Apparatus

All parts of the main study were completed on 80386 computers, with VGA displays, similar to those used in Study 1. The Armed Services Vocational Aptitude Battery (ASVAB) made up the entire ability battery. The ASVAB was completed by the trainees several months prior to the experiment and test scores were obtained from personnel records.

Ability Tests

The ASVAB has been extensively described elsewhere (e.g., see the *Technical Supplement to the Counselor's Manual for the Armed Services Vocational Aptitude Battery Form-14*, 1986). The battery contains 10 tests, as follows: General Science, Arithmetic Reasoning, Word Knowledge, Paragraph Comprehension, Numerical Operations, Coding Speed, Auto and Shop Information, Mathematics Knowledge, Mechanical Comprehension, and Electronics Information. Composite scores were derived from the original tests, in a unit-weighted z-score procedure. The composites were: Verbal, Vocational-Technical Information, Math, and Perceptual Speed abilities.

Personality Measures

In order to allow investigation of the relations between the criterion performance and the broad personality domain, a short form personality test, NEO-FFI was administered (Costa and McCrae, 1992). The NEO-FFI is a 60 item inventory designed to assess standing on five broad factors of personality. The five broad personality factors assessed by the NEO-FFI are: Extroversion, Neuroticism, Openness, Agreeableness, and Conscientiousness. Additional personality measures included: Achievement Motivation (WOFO, Helmreich & Spence, 1978), Typical Intellectual Engagement (Goff & Ackerman, 1992), Action Control (Kuhl, 1985), Functional Impulsivity (Dickman, 1990), and Control (MPQ, Tellegen, 1982).

Self-Ratings of Ability and Self-Concept

For self-ratings of ability, a 22-item questionnaire was used. The items represent three broad factors of abilities: spatial-math ability, verbal ability, and self-regulation ability. Subjects were instructed to respond with a self-evaluation *relative to other persons their age*.

Task Characteristics.

The Target/Threat task was used in a format identical to that described for Study 1, though only the CM single-task condition was used.

Procedure

The procedure in this study followed the general form of Study 1. Trainees first completed the self-report measures. After a break, trainees first completed the target/threat instructions, and then performed 30 trials of the CM condition of the target/threat task. At the conclusion of the task trials, a memory recall test was administered.

Results

Target/Threat Performance. Overall performance in the CM version of the threat task was substantially poorer, and had more widespread individual differences than in Study 1. RT $\underline{M}$ = 2,035 ms, $\underline{sd}$ = 448 ms (in comparison to Study 1 RT $\underline{M}$ = 1730, $\underline{sd}$ = 429). Similarly, memory performance at the end of CM practice was relatively poor, $\underline{M}$ = 4.67, $\underline{sd}$ = 2.8 (especially in comparison to Study 1 $\underline{M}$ = 8.85, $\underline{sd}$ = 1.5). Thirty-one of the trainees had a score of zero on the memory test, indicating that many trainees did not likely take advantage of the consistencies of identifiers in the task.

Ability - Performance Correlations. Respective correlations between ability composites and RT were: r = .07, -.14, .14, and .19 (for Verbal, Vocational/Technical Information, Math, and Perceptual Speed) -- after performance scores were reflected so that positive correlations indicate a positive association between good performance on each measure. Only the Perceptual Speed - performance correlation was statistically significant, $p < .05$. Correlations between ability composites and memory test performance all failed to reach significance.

Personality - Performance Correlations. Similar to the low associations between ability and performance, only the Extroversion scale of the NEO-FFI significantly correlated with RT the performance measure ($r = .15$). Only the Neuroticism scale of the NEO-FFI showed a marginally significant correlation with memory test performance ($r = .12$, $p = .06$).

Self-estimates of Ability - Performance Correlations. None of the self-estimates of ability significantly correlated with task performance.

Multiple Regression Model. Based on the initial correlations, a selected variable hierarchical multiple regression was performed on RT performance. The multiple correlation between the predictor measures and task performance was significant $R^2 = .127$, though not substantial. The two significant ability predictors (Vocational/Technical Information and Perceptual Speed) accounted for 6.2% of the variance, Personality (Extroversion and Conscientiousness) accounted for 4.5% of the variance, Self-estimates of ability 1% of the variance, and general self-efficacy 1% of the variance in performance. Although less variance in performance was accounted for in this study, as compared with the more-controlled laboratory environment used in Study 1, these results were generally similar in form to those found in Study 1. The major exception to these results was the significant incremental prediction of performance that was accounted for by individual differences in personality variables.

Discussion

The discussion of general findings from the field study will have to be postponed until Drs. Vickers and Hervig complete analyses of the predictor and performance measures in light of their data, specifically the mood measures and blood workups. In the absence of these analyses, the findings from our analyses point mostly to the importance of ability predictors of performance on the Target/Threat task. One additional item that does pose questions for future analysis is the role of personality measures (specifically Extroversion and Conscientiousness) in providing significant incremental validity for individual differences in performance. Should these personality measures interact with mood and physiological measures of stress-reaction, it is possible that these data may reveal important affective-physiological-cognitive processing interdependencies. In addition, and in light of Study 3 results (see below), it appears that an adequate representation of the complex of cognitive and non-cognitive predictors of individual differences in performance under stress should also take complex Perceptual Speed abilities into account. Future field study of ability and non-ability predictors should include one or more of these measures as part of the overall prediction equation.

VII. Study 3 - TRACON Laboratory Study

Background

A series of studies mapping our integrated approach for cognitive/self-regulation aptitudes to complex skill acquisition was recently completed. There are three components of this work: (a) A laboratory investigation of skill acquisition, using the TRACON simulation; (b) A field-study for cross-validation of our approach to predicting the success/failure of Federal Aviation Administration (FAA) air traffic controllers (for a 9-week screening program, and beyond); and (c) development of a combined selection/training program for the Minnesota Air Traffic Controller Training Program (MATCTC).

Over the past few years, we completed two TRACON experiments (Ackerman, 1992; Ackerman & Kanfer, 1993). The first experiment demonstrated support for the Ackerman theory of the cognitive ability determinants of complex task performance. In the second experiment, new cognitive aptitude tests were created to increase overlap between theory and assessment instruments. Specifically, tests were developed to assess: (a) Reasoning, (b) Spatial Visualization, (c) Problem Solving, (d) Perceptual Speed, (e) Spatial Time Estimation, and (f) Spatial Memory. Additional measures of self-regulatory aptitudes were also created for predicting individual differences in skill acquisition. Specifically, measures were developed to assess: (a) emotion control, (b) action control, (c) typical intellectual engagement, (d) self-regulatory processes (self-monitoring, self-evaluation, self-reactions), and several other related constructs. One hundred and twenty-two subjects participated in this 27 hour experiment. Results from this experiment replicate the basic findings of the earlier TRACON study, and also showed that the self-regulatory measures provide substantial incremental validity in predicting both global and component processes involved in acquisition of this complex task skill. The combination of cognitive and self-regulatory aptitude measures predicted performance quite well (r = .663) -- accounting for 44% of the interindividual differences variance.

The second component of this series was a cross-validation of the laboratory study to the real-world environment of FAA selection/training of air traffic controllers. A small subset of the measures developed for the TRACON experiment (about 2.5 hours of testing) and the prediction equations developed from the TRACON experiment were applied to predicting the success/failure of N = 204 air traffic controllers in the intensive 9-week selection/training program that all controllers must pass. The predictive validity of these combined cognitive/self-regulatory aptitude measures was r = .585, which translates to accounting for 34% of the variance in performance. This result is noteworthy in that it was a strong test of our lab-based measures in a long-term real-world complex task environment. Moreover, our short test battery has higher validity than the test battery developed by the Office of Personnel and Management (OPM) for the same purpose (the validity of that battery is r = .470, accounting for only 22% of the variance in performance). Finally, our integrated battery provides information for training interventions, in a context similar to the emotion control/motivation control manipulations described earlier. Implementation of these interventions may have the pragmatic effect of reducing the wash-out rate for FAA air traffic controllers, which currently runs about 40% (even after an initial aptitude screening process that eliminates about 90% of the applicants for these positions).

The third component in this series was to refine and apply the full battery of cognitive/self-regulatory aptitude measures in an operational environment (MATCTC), with follow-up over an extended period of time (initially 6 months). This part of the project was completed in early 1993. The test battery well-predicted (r = .60) failure/success in the six-month training program for air traffic controllers, and is currently being used for selection purposes.

Study 3 Overview

The third study in this project allowed us to evaluate tests in the context of a more complex knowledge-rich task (namely, the high-fidelity Terminal Radar Approach Controller - TRACON simulation of air traffic controller activities). The laboratory environment used for this study also allowed for a broader assessment of human capacities and limitations in response to specific stress situations. The aim of this study was to set the stage for later prototype battery development and implementation for identification and/or amelioration of individual differences in stress-reactivity.

The method for this study mirrors that of the previous studies: (1) Prospective assessment of individual differences measures (e.g., cognitive ability, personality, and self-regulatory skill determinants of stress-reactivity); (2) Concurrent measurement of performance under stressful task demands; and (3) Retrospective measurement of self-regulatory processes during information processing and decision-making (e.g., ongoing self-monitoring, self-evaluation, and self-reactions). However, this study involved more extensive individual differences assessment (5+ hours) and longer term skill acquisition (18 hours) on a knowledge rich task that could not be implemented in the field-study environment.

Layout of Experimental Design -- Overview of Measures

This experiment contains four major categories of measures: Distal individual differences measures, Proximal individual differences measures, Concomitant Proximal measures, and TRACON task performance measures. These measures are described briefly below, and are presented in greater detail in the Method section.

1. *Distal Measures.* Numerous distal measures were administered, in order to converge on the multiple determinants of performance in the TRACON simulation task. The measures were designed to assess individual differences cognitive/intellectual abilities, personality, vocational interests, motivational skills, and self-ratings of ability and self-concept. In addition, the Dial Reading Test and a companion test (the Directional Headings Test) were administered to evaluate whether these measures assess common variance among ability and non-ability influences on performance.

2. *Proximal Measures.* A questionnaire (Interim Questionnaire) was administered one-two days after trainees had viewed the instructions for performing TRACON, and just prior to actual task engagement. This questionnaire assessed Negative Motivation and Positive Motivation thoughts directly pertaining to the TRACON task, and several aspects of task-specific Self-Efficacy. (This questionnaire was also administered prior to each subsequent TRACON session -- see below.)

3. *Concomitant/Proximal Measures.* Two types of concomitant measures were used during TRACON practice, an Interim Questionnaire (administered just prior to the beginning of each TRACON session) and a Task Perceptions Questionnaire (administered just after each TRACON session). Both measures assessed task-specific thoughts, both retrospective and prospective. Retrospective thoughts for the Interim Questionnaire pertained to intrusive (Negative Motivation) and purposeful (Positive Motivation) during the days between TRACON sessions. For the Task Perspective Questionnaire, retrospective thoughts pertained to the 6 simulation trials in the just-completed session. Prospective thoughts pertained to TRACON task self-efficacy (both measures), and performance expectancies (Interim Questionnaire only).

4. *Task Performance Measures.* As in previous studies, the major measures used to assess performance in TRACON relate to the number of planes successfully handled (carried to final disposition) during each simulation. Overall performance is measured, as are separate performance components of handling arrival flights and overflights.

Subjects

Subjects were recruited by signup sheets and flyers distributed around the University of Minnesota. Subjects were restricted to: (a) age between 18 and 30 years, (b) normal or corrected-to-normal vision, hearing, and motor coordination; and (c) no prior experience with ATC tasks. Ninety-three participated in the experiment. Of these subjects, 42 were males, and the age level for the whole sample was: $\underline{M}$ = 20.4 years, $\underline{sd}$ = 2.75 years. Subjects were paid $125 each for the 24 hr of participation. Payment was not contingent on task or test performance.

Apparatus

The computerized self-report measures and the TRACON trials were administered on IBM PS/2 Model 70 computers, with Intel 80386/16MHz processors and IBM 8513 monitors or Compaq Deskpro computers with Intel 80386/20MHz SX processors and NEC4D monitors. Information was displayed in text mode (for the questionnaires), or in color VGA (680 H x 480 V) resolution graphics (for the TRACON trials). Input was accomplished by keyboard responses, or with a MicroTech 3-button trackball. Audio information (from TRACON) was presented binaurally over headphones to each subject (using a SoundBlaster interface). Each computer was housed on a separate table, shielded visually and aurally by large acoustic panels that enclosed each computer. For ability testing, subjects were run in groups up to 40. For TRACON task practice, subjects were run in groups up to 20.

Ability Tests

Two batteries of tests were administered in this experiment: The first battery was the Aptitude Assessment Battery (AAB) the development of which is described extensively in Ackerman & Kanfer (1993); the second battery was used to supplement the AAB, specifically to assess perceptual speed and verbal abilities. The AAB contains 9 ability tests and an 18-item measure of motivational skills, in the following order: (1) Necessary Facts; (2) Spatial Orientation; (3) Math Knowledge; (4) Spatial Analogy; (5) Problem Solving, (6) Paper Folding, (7) Verbal Test of Spatial Ability, (8) Dial Reading Test, (9) Directional Headings. (The Motivational Skills measure follows the ability tests in the administration order). The supplemental battery contained 6 tests, in the following order: (1) Vocabulary, (2) Letter/Number Substitution, (3) Controlled Associations, (4) Subtraction and Multiplication, (5) Word Beginnings, and (6) CA-2. Descriptions of the AAB tests can be found in Ackerman & Kanfer (1993). The three verbal tests and the Subtraction and Multiplication tests in the supplemental battery are from the ETS Kit (Ekstrom, et al., 1976). The Letter/Number Substitution test is a locally developed perceptual speed test (Ackerman, 1986), and the CA-2 test is a clerical checking test published by Psychological Corporation. Together, these batteries were designed to assess the following four factors (Spatial Ability, Verbal Ability, Perceptual Speed (simple), and Math Ability). Composites for each ability are computed with unit-weighted z-scores of the component tests, and an overall general ability composite is computed across all five factors.

In addition to the four standard ability factors, we also collected data on the Dial Reading Test and a similar measure, the Directional Headings Test (developed by the FAA - Civil Aeromedical Institute, Cobb & Mathews, 1972). The Directional Headings Test is a test of memory, perceptual encoding and learning, was modeled after a test designed by the FAA Civil Aeromedical Institute (see Cobb & Mathews, 1972). Subjects are given items that include a directional letter, arrow, and degree heading (e.g., S ↑ 180). They must decide the direction implied by these indicators (or in one part of the test, decide whether two of three indicators agree), or indicate that conflicting information is presented in the item. We used these two tests to identify an ability that we initially called "Perceptual Speed (complex)," which we distinguished from an ability that is identified with more traditional, simple measures of perceptual speed.

In previous predictive validity studies, both the Dial Reading Test and the Directional Headings Test have shown high validity for predicting the success of FAA air traffic controllers, and for performance on our TRACON simulation task. In addition these two tests show high intercorrelations ($\underline{r}$ = .74, reported in Ackerman & Kanfer, 1993). In fact, similar to the findings of the Army Air Force with the Dial and Table Reading Test, the our version of the Dial Reading Test and the FAA's Directional Headings Test had mathematically the highest validity for TRACON performance in a large battery of cognitive and perceptual speed ability tests. For these two measures, a unit-weighted z-score composite was computed, and below is referred to as Perceptual Speed (complex).

Personality Measures

In order to allow investigation of the relations between the criterion performance and the broad personality domain, the NEO-PIR (Costa and McCrae, 1992) was administered. The NEO-PIR is a 240 item inventory designed to assess standing on five broad factors of personality. The five broad personality factors assessed by the NEO-PIR are: Extroversion, Neuroticism, Openness, Agreeableness, and Conscientiousness.

In addition to the five broad personality factor scales, an additional measure was included that attempts to bridge the personality and cognitive ability domain -- namely 59-item Typical Intellectual Engagement Scale. Typical intellectual engagement (TIE) is a personality construct that represents an individual's aversion or attraction to tasks that are intellectually taxing, and is thus related to acculturative and purposeful development and expression of certain intellectual abilities (e.g., crystallized intelligence). Scores on this measure have been shown to have different correlations with the broad ability classes of fluid and crystallized intelligence (with higher correlations between TIE and crystallized intelligence). (Ackerman and Goff, 1994; Goff & Ackerman, 1992).

Vocational Interest Measure

Vocational interests were assessed by using American College Testing's 90-item Unisex Edition of the ATC Interest Inventory (UNIACT) (Lamb, & Prediger, 1981). The UNIACT inventory is designed to assess six different factors of vocational interests (Holland, 1959, 1985) by asking respondents to indicate their *interest* or liking for doing particular job tasks, irrespective of their competency to do them. The six scales in the battery are Realistic, Investigative, Artistic, Social, Enterprising and Conventional.

Self-Ratings of Ability and Self-Concept

For self-ratings of ability, a 31-item questionnaire was used. The items represent three broad factors of abilities: spatial-math ability, verbal ability, and self-regulation ability. Subjects were instructed to respond with a self-evaluation *relative to other persons their age*.

Self-concepts for competencies and knowledge in several specific areas were assessed with a 32-item questionnaire developed specifically for TRACON-relevant abilities and skills. The areas assessed were self-concepts of mechanical, verbal, math, spatial, clerical, and science competencies, as well as self-concepts of self-management ability and stress-resistance.

Motivational Skills

The Motivational Skills measure is an 18-item questionnaire assessing aspects of self-confidence for learning, studying, and performing in test situations (for a description of the measure see, Ackerman and Kanfer, 1993). The Motivational Skills questionnaire has been shown to provide incremental prediction of TRACON performance after abilities are partialled out.

Self-Efficacy Measures

A series of self-report questionnaires were administered before and after the TRACON simulation trials on Sessions 3 - 8. Trainees were asked to indicate their self-efficacy for components of TRACON task performance (handling arrivals and overflights), and for overall performance (aggregate planes handled, and performance relative to the population of college students). Questions were administered in an ordered fashion. In the Interim Questionnaire, questions referred to performance on the current session, for example: "I can land *xx* percent of the planes in the upcoming session" (where *xx* refers to 20, 40, 60, 80, and 100). For the Task Perceptions Questionnaire, questions pertained to only planes handled and overall performance. These questions referred to expected performance on the next session.

Proximal/Concomitant Measures

In addition to self-efficacy, several other scales were administered either just before each TRACON session, or just after each TRACON session. Prior to task performance trainees were asked about Negative Motivational thoughts (e.g., "Between the last session and today, I wished that I was done with this project") and Positive Motivation thoughts (e.g., Between the last session and today, I imagined myself making no errors on the task). Subsequent to each session, trainees were asked questions about frequency of thoughts related to Planning, Off-task, Positive Cognitions, Negative Affect and Motivation *during the current TRACON session*.

Terminal Radar Approach Control (TRACON). The task is described in Appendix B.

Procedure

The experiment was completed in 8 sessions (one ability testing session, one session with self-report measures and a training video, and six, 3-hr TRACON practice sessions), spread out over a two week period. The first session was devoted to ability testing. The second session was devoted to personality and other self-report testing, and to viewing an instructional videotape specifically designed for the TRACON task. The videotape, 60 min in length, described the major task components, the rules regarding operation of the computer interface (keyboard and trackball), display characteristics, and procedures for accomplishing the controller task. Two "quizzes" were administered, one about halfway through the videotape, and the other at the end of the tape. Actual TRACON task practice started on Session 3, and proceeded through Session 8. Each of these practice sessions included five simulation scenarios. Each scenario (trial) lasted 30 min. Order of ability testing was constant across subjects, but order of simulations was counterbalanced in a Latin square design. Breaks were given after each simulation/task trial, and self-report questionnaires were administered at the beginning and end of each TRACON practice session. Subjects were then debriefed and paid. The entire experiment was completed in 24 hr, including 6 hr of testing, one hr of task-based video instruction, two hr of accumulated breaks, and 15 hr of total time-on-task in TRACON.

Results

Ability Tests. Means, standard deviations, reliability estimates, and intercorrelations are provided for the 15 ability tests in Table 2. These tests all performed as expected, in that they had acceptable reliabilities, and showed convergent and discriminant validity (confirmed with factor analyses). Composites were formed by unit weighted *z*-scores of the tests in each content set (i.e., Spatial Ability, Verbal, Perceptual Speed(complex), Perceptual Speed(simple), Math), with an overall General ability unit-weighted composite based on all 14 measures.

Skill-Acquisition -- TRACON Data. The purpose of this section is to describe the effects of practice on the TRACON task. First, the results are presented for overall performance. These results are followed by the results for two important components of overall performance, namely the handling of Arrivals and Overflights.

Overall performance. Previous studies with this version of TRACON have clearly demonstrated that the task is complex, and, within the parameters of our simulations, quite difficult early in practice (Ackerman, 1992; Ackerman & Kanfer, 1993). On Trial 1, the number of successful plane handles was $M = 5.89$, $sd = 4.67$. Although we consider this to be good initial performance, we should note, however, that 12% of the trainees had zero successful plane handles on Trial 1 (even though all trainees had already received an explicit one-hour video instruction on the task). By the end of 30 trials of practice, though overall performance was much improved, $M = 19.49$, $sd = 6.57$. At the end of training, 2% subjects were able to handle all 28 of the planes. Mean performance levels across practice are shown in Figure 10. A repeated measure analysis of variance (ANOVA) on overall performance confirmed the significant effect of practice that is obvious in the figure, $F(29, 2668) = 145.03$, $p < .01$.

Component performance. Although the simulations had different numbers of arrivals (12/trial) and overflights (approximately 8/trial), examination of performance on each type of flight helps illustrate the difference in difficulty levels of the respective types of flights. Mean performance levels for arrivals and overflights are shown in lower portion of Figure 10. For Trial 1, $M_{\text{arrivals}} = 1.69$, $M_{\text{overflights}} = 2.51$. At the end of practice, substantial improvement on both types of flights was indicated -- Trial 30, $M_{\text{arrivals}} = 7.99$, $M_{\text{overflights}} = 6.51$. Although trainees were handling a greater number of arrivals than overflights at the end of practice, it is important to keep in mind the different number of flight types available in each trial. That is, at the end of practice, mean performance reflected the fact that trainees were handling 66% of the available arrivals, but 81% of available overflights. (To save space, and because departures [approximately 8/trial] represent moderate difficulty -- i.e., more difficult than overflights, but less difficult than arrivals, these component measures are not included.)

Table 2. Pearson Correlation Matrix

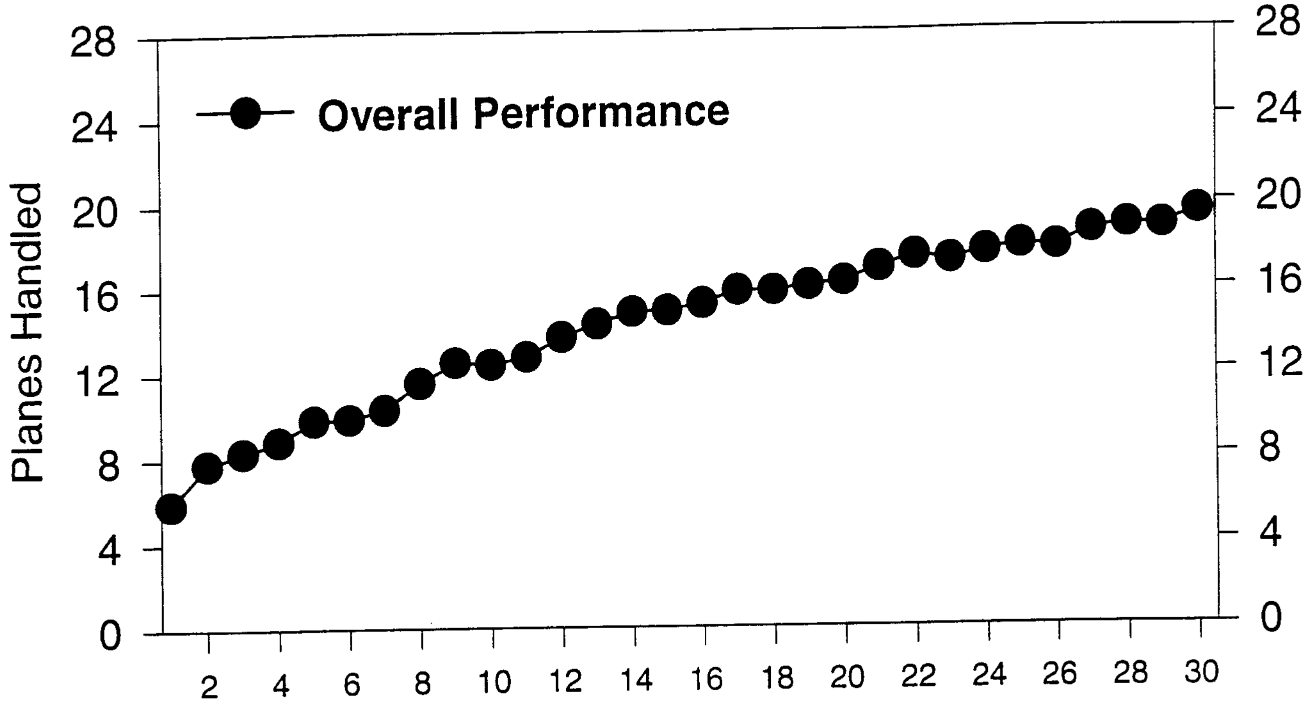
	M	SD	r_{xx}	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Necessary Facts	7.92	4.18	.74														
2. Spatial Orientation	7.13	3.66	.53	.250													
3. Math Knowledge	14.86	8.48	.72[a]	.571	.158												
4. Spatial Analogy	15.72	5.04	.92[a]	.558	.323	.404											
5. Problem Solving	5.12	2.70	.85[a]	.595	.300	.555	.412										
6. Paper Folding	12.78	4.49	.72	.574	.437	.433	.585	.523									
7. Verbal Test of Spatial Ability	9.88	3.91	.60[a]	.569	.429	.407	.573	.501	.578								
8. Dial Reading Test	88.81	32.60	.94	.582	.133	.460	.500	.571	.482	.458							
9. Directional Headings	63.82	18.42	.84	.498	.175	.463	.580	.583	.464	.536	.776						
10. Vocabulary	21.01	8.55	.86	.303	-.018	.130	.178	.398	.241	.412	.430	.381					
11. Letter/Number Substitution	131.48	21.54	.92	.376	.120	.180	.431	.368	.388	.421	.486	.484	.216				
12. Controlled Associations	21.77	7.00	.67	.354	-.014	.330	.259	.340	.304	.451	.503	.409	.643	.221			
13. Subtraction and Multiplication	48.86	18.08	.90	.299	.034	.299	.049	.404	.162	.314	.483	.380	.278	.352	.349		
14. Word Beginnings	17.19	6.97	.51	.119	-.092	.109	-.053	.292	.065	.251	.247	.192	.501	.157	.331	.266	
15. CA-2	35.85	8.75	.__[b]	.306	.101	.242	.297	.431	.302	.376	.658	.622	.402	.568	.408	.580	.313

Note: Correlations greater than r = .201 are significant at p = .05. M = mean, SD = standard deviation.

Reliability based on Part 1/Part 2 Correlations, corrected for test length with the Spearman-Brown prophecy formula.

[a] One part test. Indicates Kuder-Richardson formula 20 reliability.

[b] One part speeded test. Reliability calculation not possible.



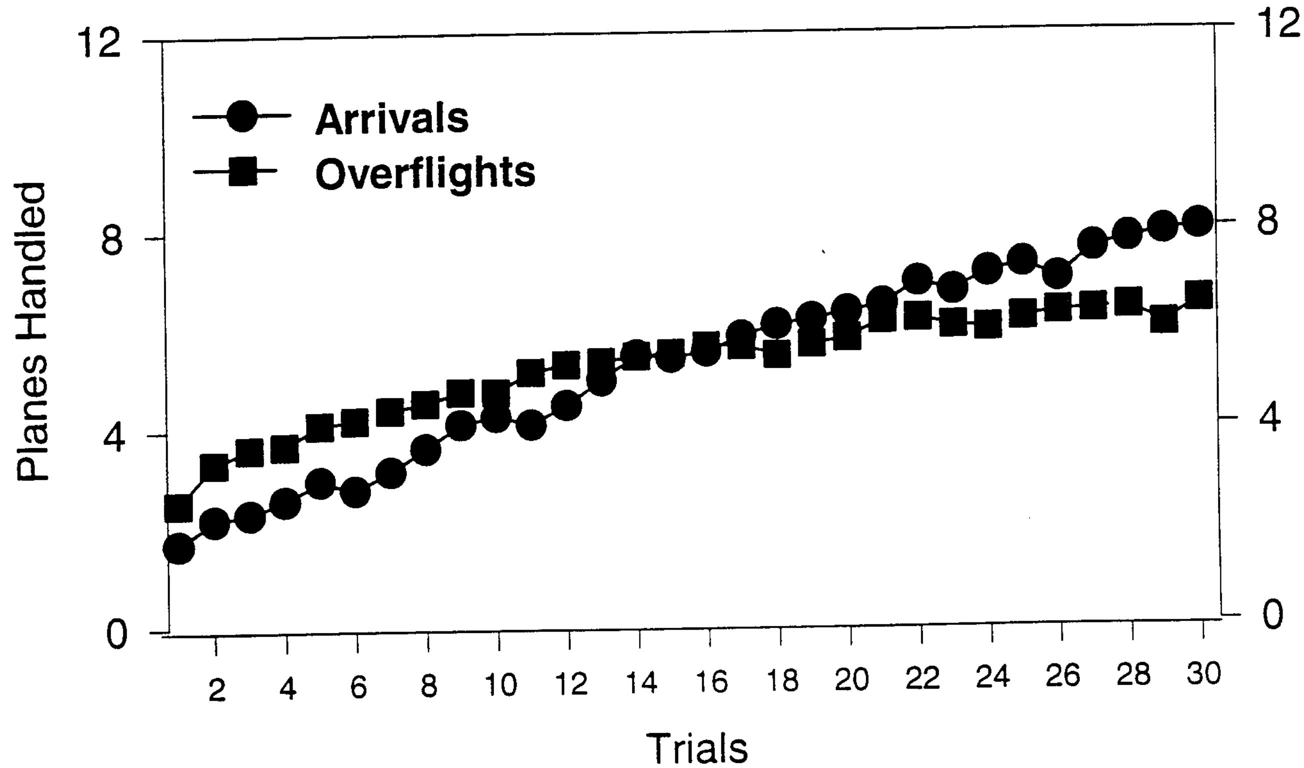


Figure 10. Top panel: TRACON performance -- mean total planes handled as a function of trials of practice. Bottom panel: TRACON performance -- mean arrivals and overflights handled as a function of trials of practice.

Ability-Performance Relations

Although the ability requirements of the TRACON task have been described elsewhere (Ackerman, 1992; Kanfer & Ackerman, 1993), for a discussion of interrelations among performance predictors it is necessary to fully document the substantial cognitive demands of the task, when trainees first confront the task, and again after substantial task practice has occurred. For these and subsequent analyses, we have concentrated on macro changes in performance, by averaging scores on TRACON by session of practice (5 simulation trials were administered on each of the 6 practice sessions). In a task that is much more quickly learned -- such as the Target-Threat task, such large time slices obscure the learning process. However, for a complex task such as TRACON, this level of analysis provides additional stability for the performance scores, and more manageable data for structural analysis, without sacrificing validity or obscuring the patterns of learning and skill acquisition. The analyses focus on overall TRACON performance, as well as differences between performance components of handling arrivals and overflights. The first section will describe the correlations between the ability composites and TRACON performance over practice.

Overall Performance. Overall performance is well predicted across all sessions of practice by the general ability composite, as shown in Figure 11, with average correlation of r = .57. Examination of constituent content ability measures, shown in Figure 12 indicates that only the Math and Spatial broad content abilities are substantial (mean r = .56 and .53 respectively), and stable predictors of performance, the Verbal content ability shows only a modest correlation with performance, mean r = .24. The two Perceptual Speed abilities show substantial validity in predicting overall TRACON performance, with the PS-Complex (mean r = .60) showing higher validity than PS-simple (mean r = .44). These results are consistent with previous task analyses and extant validity data in our previous two studies (Ackerman, 1992; Ackerman & Kanfer, 1993).

Secondary analyses can be predicated on general performance, across all sessions of practice. Toward this end, a single planes-handled measure was computed across all 30 trials, and correlated with the general and the constituent content and perceptual speed ability composites. The correlations between abilities and overall performance are as follows: General r = .59, Spatial r = .56, Math r = .59, Verbal, r = .24, Perceptual Speed (complex) r = .63, Perceptual Speed (simple) r = .46 (all significant $p < .05$). For comparison, a single-step multiple regression (with all ability measures except for the aggregate general composite) yielded a multiple R of .70, or an $R^2 = .49$. That is, taken together, the five ability composites accounted for 49% of the variance in overall TRACON performance, across 15 hours of task practice.

Arrivals and Overflights. Separating the performance components of arrivals and overflights illustrates the task-analysis that described arrivals as more difficult and more demanding of spatial abilities than overflights (which are hypothesized to be more easily learned, and thus more highly associated with perceptual speed abilities). Figure 13 shows that initial performances on both arrivals and overflights are equivalently well-predicted by

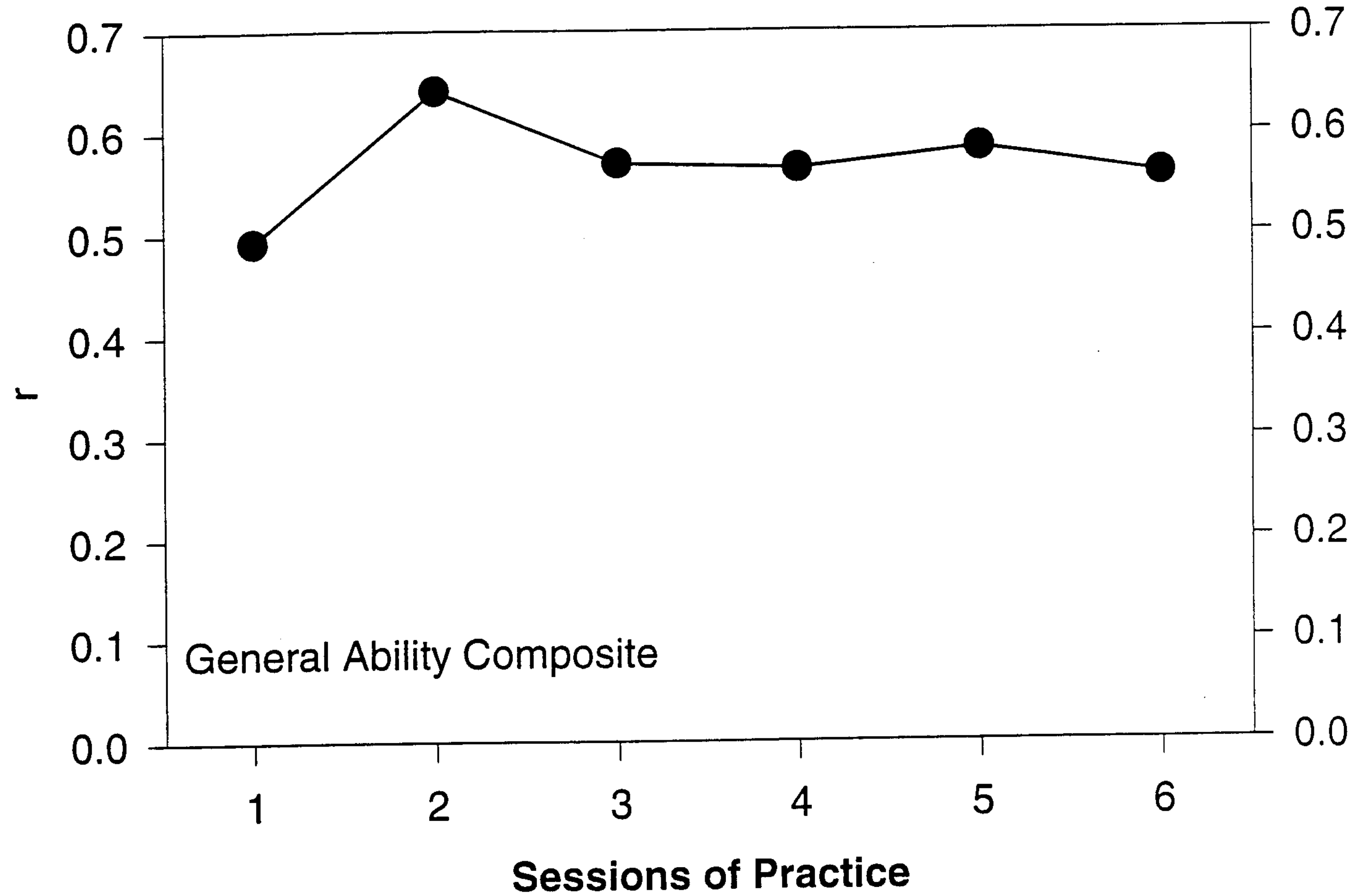


Figure 11. Correlations between general ability composite and overall performance (planes handled) on TRACON, over sessions of practice (5 simulation trials per session).

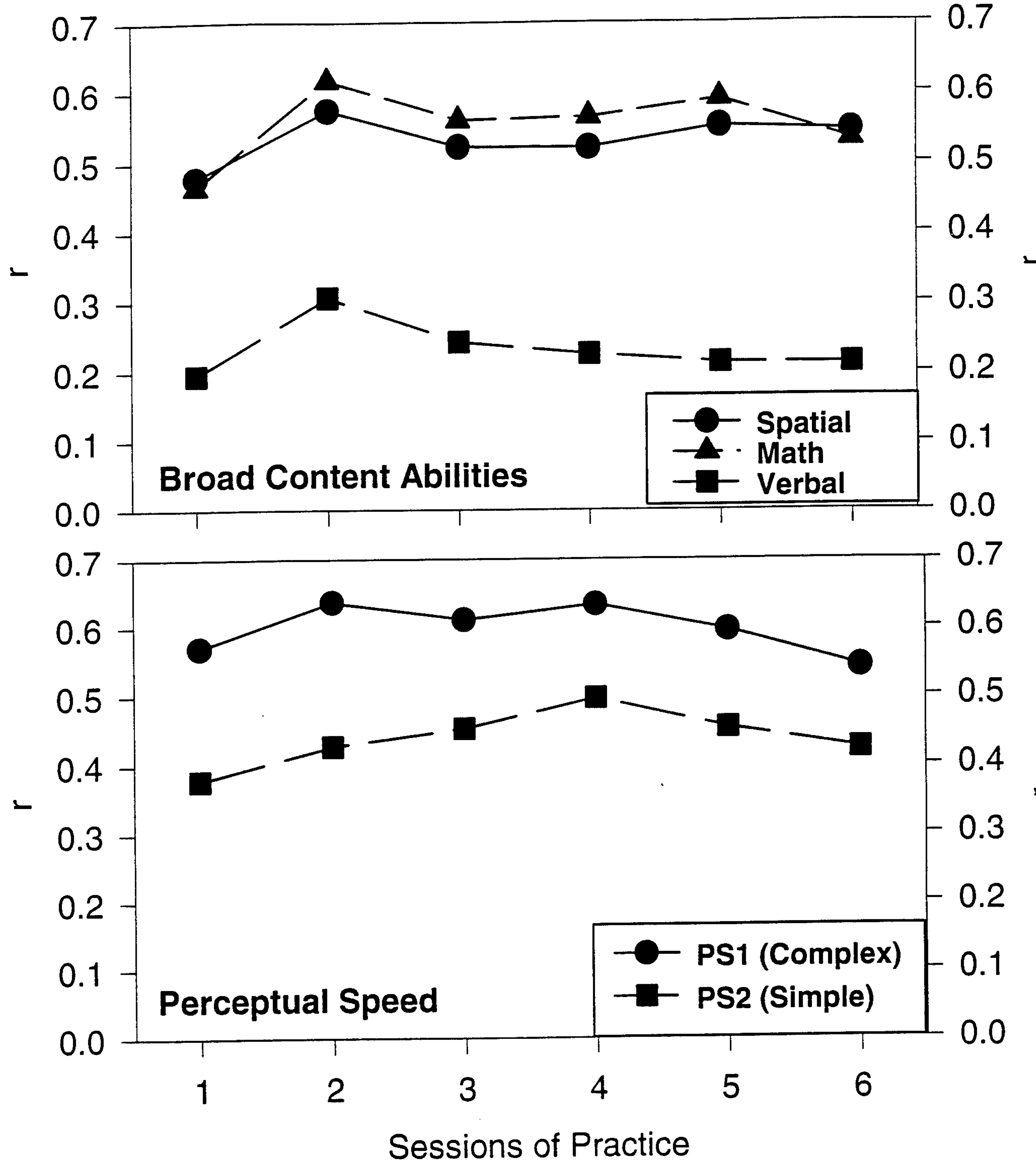


Figure 12. Correlations between ability composites and overall performance (planes handled) on TRACON, over sessions of practice (5 simulation trials per session). Top panel: Content ability composites (spatial, math, and verbal). Bottom panel: Perceptual Speed (complex) and Perceptual Speed (simple).

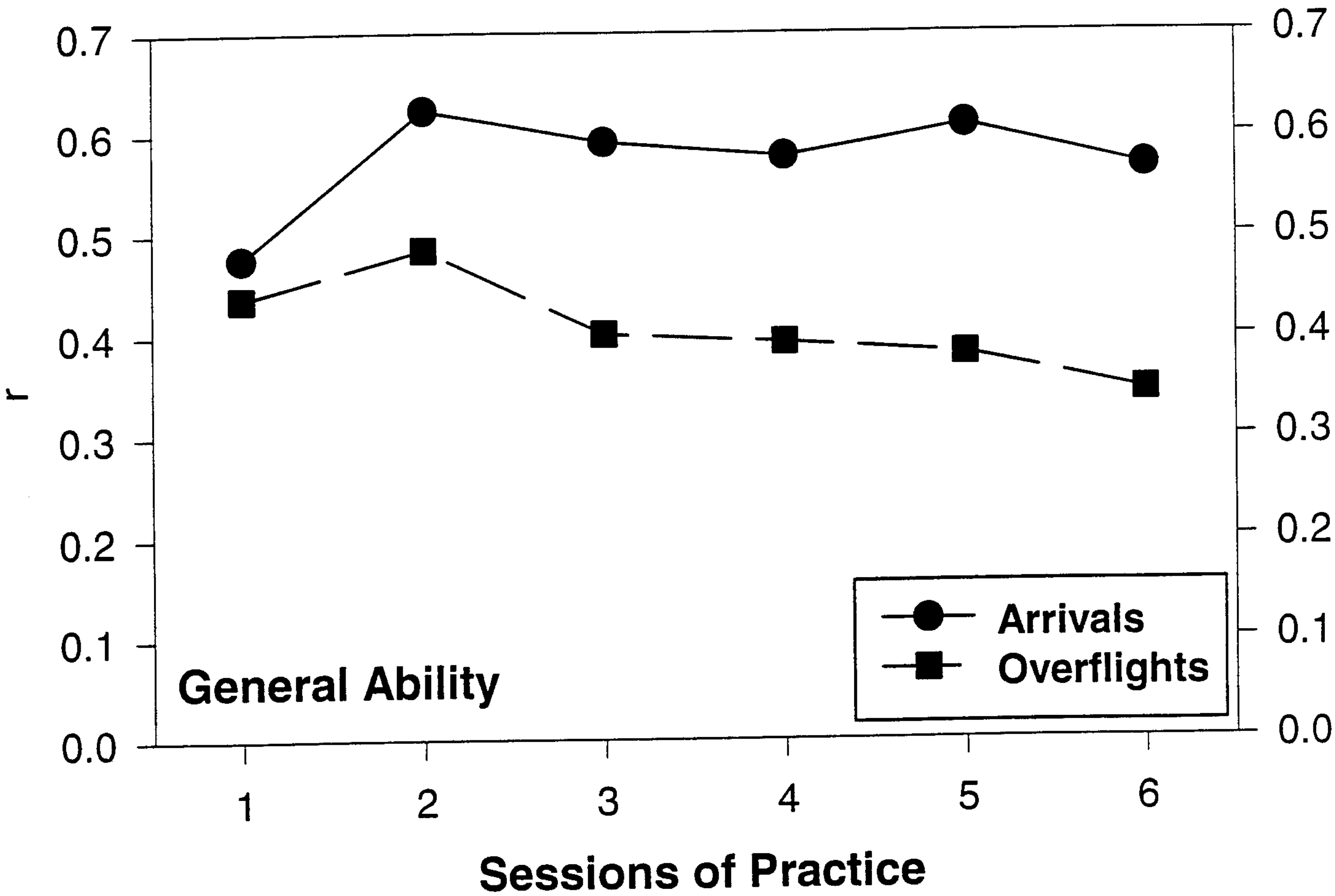


Figure 13. Correlations between general ability composite and performance components (arrivals and overflights) on TRACON, over sessions of practice (5 simulation trials per session).

the general ability composite. As predicted though, with practice, validity coefficients for arrivals and overflights diverge, with general ability showing higher correlations with arrival performance, and lower correlations with overflights.

Figure 14 shows arrival and overflight correlations for four ability composites (Verbal ability having been dropped, as a result of demonstrated lack of substantial validity for predicting TRACON performance). Again, concordant with our expectations, Spatial ability and Math ability show divergence in validity for predicting arrivals after initial practice (with greater validity for predicting arrivals), and the two perceptual speed measures show greater similarity in predicting performance of both arrivals and overflights.

Summary and Discussion. The results described above support several points about the relations between abilities and performance on TRACON, as follows:

1. Performance throughout practice is well-predicted by measures of cognitive and intellectual ability. The measures most highly associated with task performance were Spatial, Math and Perceptual Speed (complex) abilities. Verbal ability showed only a modest validity for predicting performance on TRACON, at any stage of practice.

2. Across practice, ability measures accounted for nearly 50% of the variance in TRACON performance.

3. Arrival and Overflight components of performance showed substantial and equivalent ability demands early in practice, but with practice, arrivals were more highly associated with spatial and math abilities, and overflights more highly associated with perceptual speed abilities.

Gender Differences

A robust finding in the ability literature is group differences in performance on tasks that require complex spatial information processing, with males showing an overall performance advantage over females (e.g., Lohman, 1986). Our previous studies have demonstrated similar group differences in performance on TRACON, with male trainees showing a substantial performance advantage relative to female trainees. Because some of these differences may be determined by ability differences, by personality and other non-ability differences, it is necessary to review the basic performance data for TRACON by gender.

TRACON performance. Similar to previous studies (e.g., Ackerman, 1992), there was both a significant main effect for trainee gender on overall performance $F(1, 91) = 19.66$, $p < .01$), and an interaction between trainee gender and practice $F(5, 455) = 5.49$, $p < .01$). The means for each group over practice are shown in the top panel of Figure 15. As can be seen in the figure, the interaction is indicated as a divergence in performance between the two groups with increasing practice on TRACON.

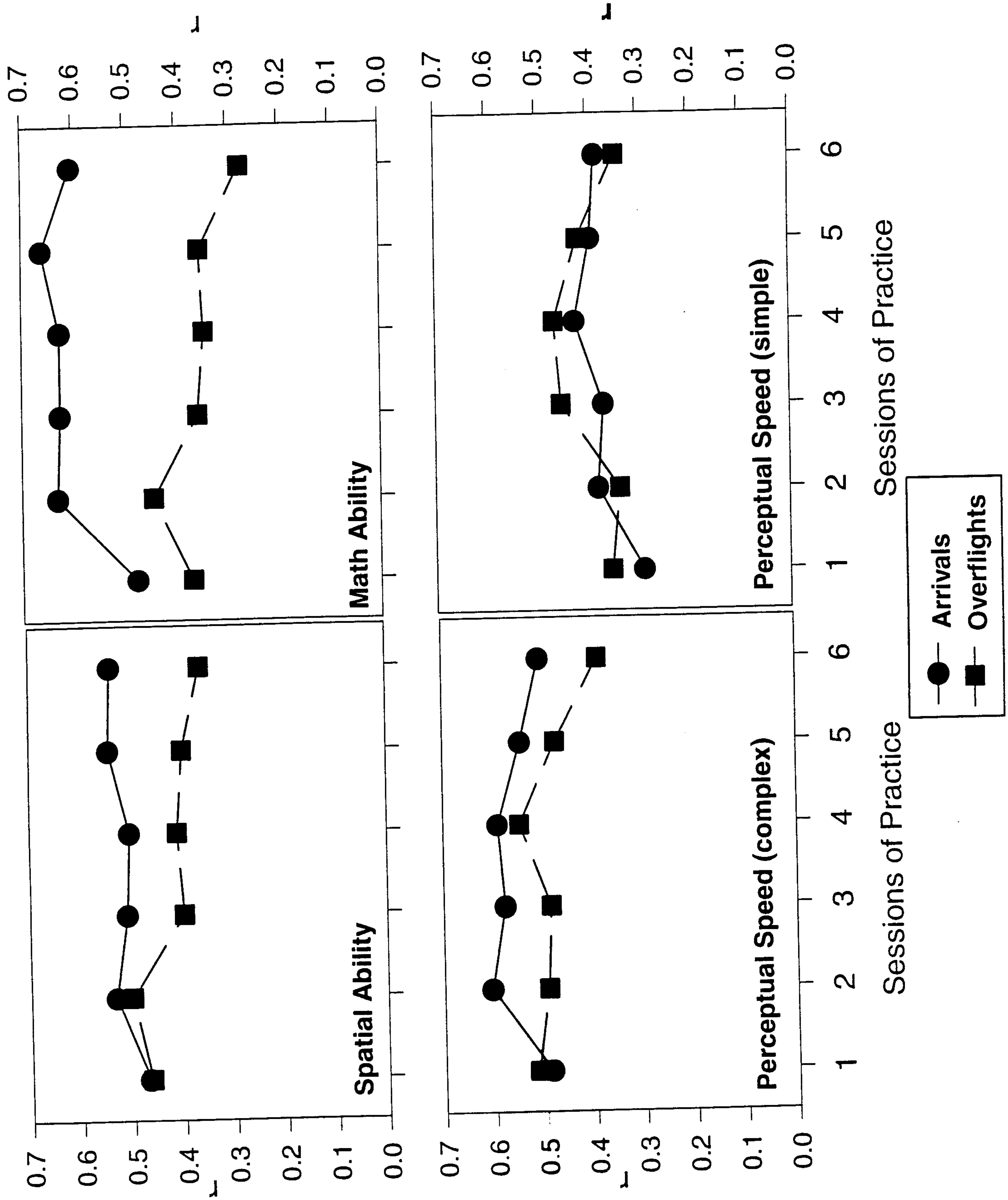


Figure 14. Correlations between ability composites and performance components (arrivals and overflights) on TRACON, over sessions of practice (5 simulation trials per session).

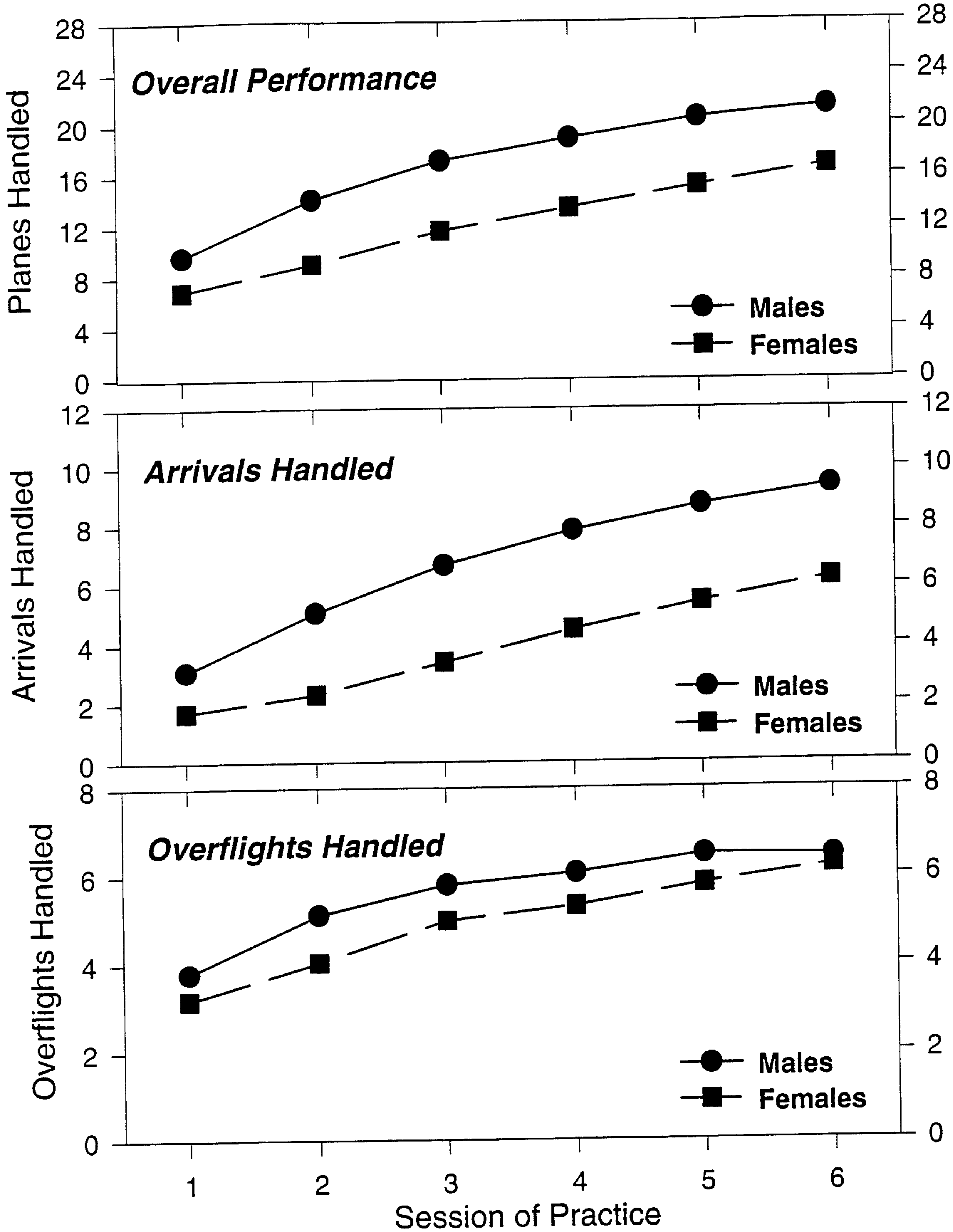


Figure 15. Mean TRACON performance over sessions of practice, by gender. Top panel: Overall performance. Middle panel: Arrivals. Bottom panel: Overflights.

Concordant with the earlier ability-performance analyses of arrivals and overflights, which showed higher correlations between arrival performance and math/spatial abilities than for overflights, the differences between gender groups show greater advantages to males over females in handling arrivals. The lower panels of Figure 15 show these effects. A repeated measure ANOVA of arrivals showed a main effect of gender F (1, 91) = 32.25, $p < .01$, and a significant interaction between gender and practice F (5, 455) = 7.74, $p < .01$), with performance diverging between groups with practice. In contrast, a repeated measure ANOVA on overflights showed a significant, but smaller, effect of gender F (1,91) = 4.32, $p < .05$, and a nonsignificant interaction between gender and practice F (5, 455) = 2.05, ns. At the end of practice, males and females have equivalent performance on overflights ($M_{\text{males}} = 6.44$, $M_{\text{females}} = 6.21$). Again, these results support the assertion that handling arrivals requires more complex spatial information processing than does handling overflights.

Non-ability predictors

Personality. Personality constructs have often been implicated in determinations of stress-reactivity or stress-resistance (e.g., see Vickers, 1991). As such, we first correlated measures of the five major personality composites with overall performance on TRACON. None of the correlations reached traditional levels of significance: Neuroticism $r = -.15$, Extroversion $r = -.02$, Openness $r = -.01$, Agreeableness $r = .02$, and Conscientiousness $r = .02$. A multiple regression with these five factors similarly yielded no significant result, $R = .17$, or $R^2 = .03$. Simple correlations between personality composites and daily session measures of TRACON yielded similar results, as did separate correlations for arrivals and overflights.

Vocational Interests. It has often been claimed in the vocational literature that emotional stress is generated when a mismatch exists between an individual's vocational interests and the task that the individual is asked to perform (see, e.g., Dawis & Lofquist, 1984). As such, we computed correlations between theme-scales of the UNIACT and aggregate TRACON performance, with the following results: Realistic $r = .31$, Investigative $r = .19$, Artistic $r = -.14$, Social $r = -.07$, Enterprising $r = -.07$, and Concrete $r = .07$. The correlation between the Realistic theme-scale and performance is the only one that reaches both statistical and practical significance, accounting for about 10% of the variance in TRACON performance.

Self-Ratings of Ability and Self-Concept. The three composite scales of ability self-ratings and the eight measures of academic/ability self-concept were factor analyzed to yield three main composites. The first composite (SR1) was comprised of items from math- spatial- mechanical- and science- self-concept, and self-ratings of similar abilities. The second composite (SR2) was comprised of items from verbal self-concept and self-ratings of general verbal ability and more specific reading, vocabulary, and writing abilities. The third composite (SR3) was comprised of items from self-management- clerical- and stress-resistant self-concept, and self-ratings of self-control and coping abilities. Correlations between these three composites and TRACON performance indicated that the first composite was highly and consistently related to performance (mean $r = .46$), and the remaining two composites were essentially unrelated to TRACON performance (see Figure 16). These results indicate that selected self-ratings of ability and academic/ability self-concept are effective predictors of performance in a complex task. Those ratings (specifically of math, spatial, and mechanical abilities), accounted for about 20% of the variance in TRACON performance across practice sessions.

Motivational Skills. As in previous studies (Ackerman & Kanfer, 1993), an 18-item measure of Motivational Skills was included in this study. Consistent with results from predicting success in air traffic control tasks (both in the laboratory and in the field with FAA trainees), the Motivational Skills measure showed consistent, modest and significant correlations with performance across all sessions of TRACON performance (mean $r = .25$) -- in this case accounting for about 6% of the variance in TRACON performance.

Proximal Measures

In addition to the sets of distal individual differences measures administered prior to trainees' exposure to TRACON, an additional set of proximal (and concomitant) measures was administered. Specifically, prior to each session of TRACON practice, the Interim Questionnaire (16 items) was completed by the trainees. Similarly, at the end of each session of TRACON practice, a Task Perceptions Questionnaire (44 items) was completed by the trainees. Of critical importance is the fact that the first Interim Questionnaire was administered *prior to any direct TRACON experience (but one-two days after the video instruction on TRACON was presented to the trainees)*. As such, the first interim questionnaire captures the proximal anticipations of trainees for their initial confrontation with TRACON.

Interim Questionnaires. Two subscales from the Interim Questionnaires are of critical importance for the current investigation, namely: Negative Motivation and Positive Motivation. The Negative Motivation scale is composed of 6 items such as: "I wished that I was done with this project" and "I got discouraged when I thought about today's session." The Positive Motivation scale is composed of 7 items such as: "I imagined myself doing extremely well on today's session" and "I made a specific plan for how to perform some part of the task." Both measures significantly predicted initial performance on TRACON (Session 1), Negative Motivation $r = -.42$, and Positive Motivation $r = .22$. In addition, the administrations of these items prior to task engagement showed persistence in predicting TRACON performance over the subsequent days of task practice (see Figure 17). The panels of the figure show two curves for each measure. The first curve is the predictive validity of the Day 1 scale score throughout TRACON practice. The second curve is the predictive validity for each measure administered just prior to that day's task performance (e.g., Day 2 Interim Questionnaire with Day 2 performance). *As the results for both Negative Motivation and Positive Motivation indicate, trainees' thoughts prior to any actual TRACON experience better predict later TRACON performance than do measures administered prior to each day's session.*

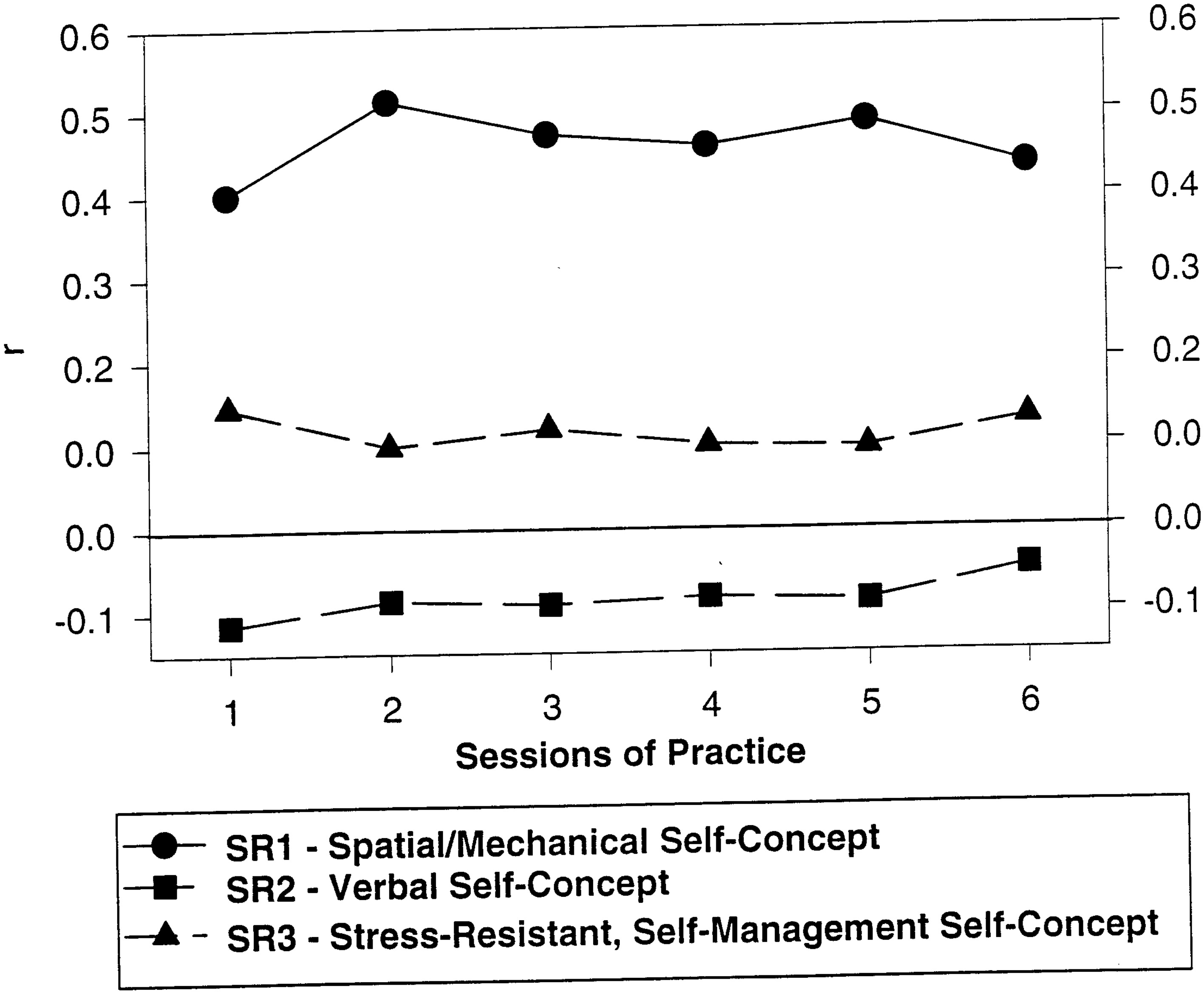


Figure 16. Correlations between composite measures of self-concept and self-estimates of ability and TRACON performance over sessions of practice.

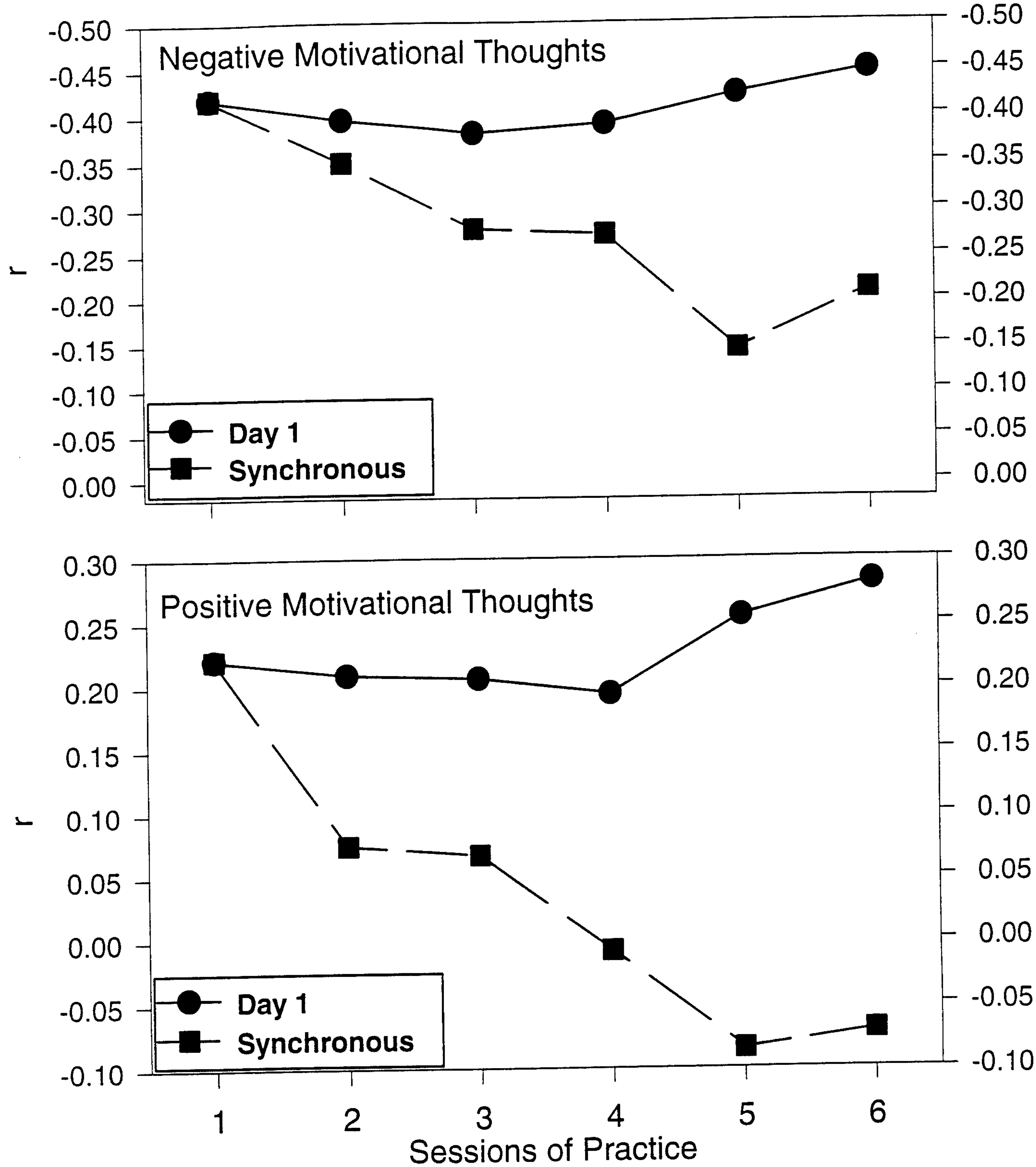


Figure 17. Correlations between self-reports of task-relevant thoughts and TRACON performance over sessions of practice. Day 1 self-reports were administered prior to actual TRACON practice. Later "synchronous" measures prior to each subsequent day's session. Top Panel: Negative Motivational Thoughts. Bottom Panel: Positive Motivational Thoughts.

Self-Efficacy (Prior to Task Engagement). At the end of the Interim Questionnaire, trainees were asked to indicate their self-efficacy for components of TRACON task performance (handling arrivals and overflights), and for overall performance (aggregate planes handled, and performance relative to the population of college students). Note that, as with the Negative Motivation and Positive Motivation items, these measures were administered two days after trainees had viewed the one-hour instructional video on TRACON, but just before the trainees had any direct experience in performing TRACON.

All four of the self-efficacy measures correlated significantly with initial TRACON performance ($SE_{arrivals}$ - $r = .23$, $SE_{overflights}$ $r = .32$, $SE_{planes\ handled}$ $r = .26$, and $SE_{performance}$ $r = .38$.) These measures were also highly intercorrelated, so that an aggregate measure of self-efficacy correlated $r = .32$ with initial TRACON performance. These correlations increased after the first session, and were stable throughout the remaining sessions. For example, the composite self-efficacy measure correlated $r = .40$, .38, .39, .42, and .43 with performance on the respective five remaining sessions of TRACON performance. The self-efficacy measure that showed the highest individual validity was the "performance" measure -- it correlated $r = .48$ with overall performance on TRACON, accounting for 23% of the variance in performance across practice sessions.

In summary, measures of proximal thoughts and self-efficacy prior to engagement in TRACON performance showed validity for predicting individual differences in task performance. Specifically, greater incidence of Negative Motivation thoughts was significantly and substantially associated with lower TRACON performance throughout practice, and to a somewhat smaller degree, greater incidence of Positive Motivation thoughts was significantly associated with positive TRACON performance throughout practice. Also, higher levels of self-efficacy (especially as perceived in comparison with other college students), were associated with both initial and subsequent TRACON performance.

Measures taken during skill acquisition

Negative Motivation and Positive Motivation. Repeated measure ANOVAs on the Negative Motivation and Positive Motivation scales revealed significant increases for both Negative Motivation thoughts ($F(5,455) = 4.02$) and Positive Motivation thoughts ($F(5,455) = 9.92$) over Sessions of practice (both $ps < .01$). However, the Negative Motivation scale showed a main effect for Gender ($F(1,91) = 3.95$, $p < .01$), with females showing overall higher reporting negative motivation thoughts, and a significant interaction between Gender and Session ($F(5,455) = 3.08$, $p < .01$), which reflected a convergence of male and female reporting of negative motivation thoughts over practice. No gender main or interaction effects were found for Positive Motivation.

Incidence of Negative Motivation or Positive Motivation thoughts, subsequent to the initial TRACON practice session showed diminished validity for predicting TRACON performance. In the aggregate, these results suggest that trainees' initial metamotivational strategies for task learning have important and persistent influences on task performance, but subsequent experience with the task (which raised the frequency of both types of thoughts) attenuates the validity of such individual differences measures for predicting performance. Such effects are consistent with the idea that trainees who are most motivationally reactive (either negatively or positively) to an upcoming stressful learning and performance experience are most likely to show persistent effects over training. In contrast, direct experience with the stressful task has more generic effects on all trainees, which eclipse differences in anticipatory motivational processes.

Task Perceptions Questionnaires. In contrast to the Interim Questionnaires (which were administered prior to each TRACON session), Task Perceptions Questionnaires were administered immediately following the last simulation for each of the six TRACON sessions. These measures concerned the frequency of thoughts that trainees had during the TRACON session. Measures of Planning, Off-task, Positive Cognitions, Negative Affect and Motivation were included. Repeated measures ANOVAs were performed for each of these measures, with Gender as a between-subject factor and Practice as a within-subject factor. Results of these ANOVAs are presented in Table 3. Results indicated several significant differences between gender groups, differences over practice, and interactions between two factors.

For frequency of Planning thoughts, there was no main effect of gender, but significant increases in planning over the course of practice, and the interaction of gender by practice. Examination of the means showed that the interaction was the result of females reporting less planning than males early in task practice, but more planning later in task practice.

For off-task thoughts and negative affect thoughts, only a main effect of practice was significant. The means showed that while Session 1 off-task and negative affect thought frequency was high (at the point where performance was lowest), the frequency dropped substantially on Session 2. Off-task thoughts showed a slow increase through later sessions of practice, while negative affect thoughts remained stable after Session 2. These results are consistent with earlier studies (e.g., Kanfer & Ackerman, 1989), that indicate off-task thoughts are often indicative of problems in negative affective responses to stressful task demands, that is, emotion control problems.

Interim Self-Efficacy Measures. Four measures of self-efficacy for TRACON performance were administered prior to each TRACON session. These pertained to self-efficacy for: (1) Percent of Planes Handled, (2) Performance (relative to other students), (3) Overflights, and (4) Arrivals. Mean self-efficacy scores for each of the four measures over practice by gender are presented in Figure 18. Repeated measure ANOVAs for each are shown in Table 4. For each of the four self-efficacy measures, there were substantial and significant main effects of Gender on scores (with females having lower mean self-efficacy for all aspects of the TRACON task), and Sessions of Practice (with self-efficacy dropping from Session 1 to Session 2, and then rising after Session 2). Significant Gender X Sessions of Practice interactions were found for $SE_{Handled}$ and $SE_{Arrivals}$, with a divergence of scores

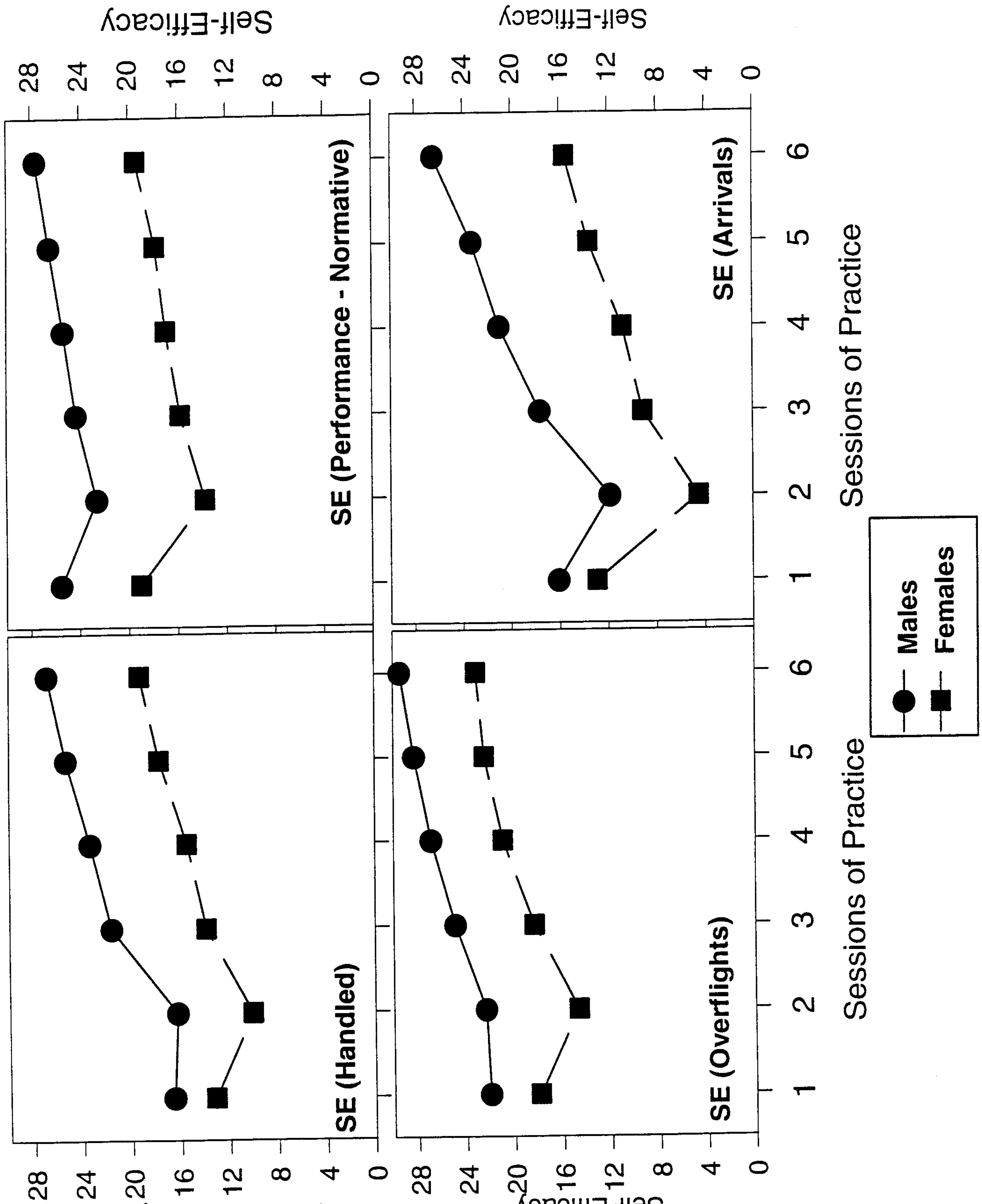


Figure 18. Mean Self-Efficacy expectation composite scores over TRACON practice, by gender. Measures administered *prior* to each TRACON session.

Table 3. Repeated-measures ANOVA for Task Perceptions Measures over TRACON Practice Sessions

	Task Perceptions Measures								
	Planning		**Off-task**		**Positive Cognitions**		**Negative Affect**		
df	MS_e	*F*	MS_e	*F*	MS_e	*F*	MS_e	*F*	
Gender	1,90	669.46	.20	379.06	.43	1005.61	.47	820.87	1.02
Sessions of Practice	5,450	43.52	14.12**	44.94	13.92**	65.19	.78	71.83	22.49**
Gender X Sessions of Practice	5,455	43.52	3.14**	44.94	1.25	65.19	1.47	71.83	.43

	Motivation		
df	MS_e	*F*	
Gender	1,90	137.80	.00
Sessions of Practice	5,450	9.14	2.02
Gender X Sessions of Practice	5,450	9.14	.33

Note. * $p < .05$; ** $p < .01$.
Mean square errors are given in parentheses.

Table 4. Repeated-measures ANOVA for Self-Efficacy Measures over TRACON Practice Sessions

	Self-Efficacy Measures								
		SE_{Handle}		$SE_{Performance}$		$SE_{Overflights}$		$SE_{Arrivals}$	
df	MS_e	*F*	MS_e	*F*	MS_e	*F*	MS_e	*F*	
<u>Interim Questionnaire</u>									
Gender	1,91	214.66	29.14**	289.46	31.92**	331.36	15.39**	292.92	32.46**
Sessions of Practice	5,455	21.70	61.73**	23.64	13.92**	27.37	32.02**	25.91	69.58**
Gender X Sessions of Practice	5,455	21.70	3.23**	23.64	.70	27.37	1.14	25.91	7.06**
<u>Task Perceptions Questionnaire</u>									
Gender	1,91	317.59	33.13**	258.28	30.29**				
Sessions of Practice	5,455	18.48	26.08**	21.26	99.33**				
Gender X Sessions of Practice	5,455	18.48	.74	21.26	1.74				

<u>Note</u>. * p < .05; ** p < .01.
Mean square errors are given in parentheses.

for males and females, and no significant interactions for $SE_{Performance}$ or $SE_{Overflights}$. It is interesting to note that even after five sessions of TRACON practice, the final $SE_{Performance}$ and $SE_{Arrivals}$ ratings by females did not exceed those of males *at Session 1*, even though mean overall performance on TRACON showed a substantial advantage by females at Session 6 vs. males at Session 1.

Task Perceptions Self-Efficacy Measures. Two measures of self-efficacy for TRACON performance were administered immediately after each TRACON session. These pertained to self-efficacy for: (1) Percent of Planes Handled and (2) Performance (relative to other students), in a format that was identical to the Interim Questionnaire measures. Mean self-efficacy scores for each of the four measures over practice by gender are presented in Figure 19. Repeated measure ANOVAs for each are shown in the bottom of Table 4. These measures show similar patterns to those from the Interim Questionnaires, with the exception of the absence of a drop in scores from Session 1 to Session 2. The obvious reason for this difference is that the Session 1 self-efficacy measures were administered after the trainees had their initial experience with TRACON, and the first Interim Questionnaire was administered prior to actual performance experience. Significant main effects for Gender and for Sessions of Practice were found for both measures, but no significant interactions were found for either measure. Such results indicate general improvement in self-efficacy with practice, and persistently higher self-efficacy scores for males.

Multiple Regression/Correlation Results.

As in the previous studies, raw correlations between various predictors and task performance only give partial information, in that each set of predictors is treated in isolation. Common variance among predictor variables serves to limit the overall predictive validity of an integrated battery of tests. Multiple correlation/regression procedures allow for the determination of the incremental validity of each set of predictor variables in a general equation for predicting task performance. Although the battery of tests and questionnaires was larger in Study 3 than in the previous studies, the general logic of the analyses we followed was identical. That is, (distal) ability measures were entered into a hierarchical regression first, then measures of motivational skills, self-estimates of ability and self-concept, and interests. After the distal measures were entered, proximal measures were included in the equation to examine any incremental validity from these measures. As with the previous studies, only measures taken prior to task performance were included in the prediction equation. Finally, gender was added to the equation to examine whether gender differences accounted for variance in performance, *after all the predictor measures had been taken into account.* Table 5 presents these analyses of first and last session performance for overall performance, and separately for overflights and arrivals.

Prediction of individual differences in performance was quite good. The predictors accounted for roughly half of the variance in performance, though the measures were better predictors of the arrival task component than the overflight task component (possibly because of ceiling effects on overflights). Consistent with the form of results found for the Target/Threat task in Study 1 and Study 2, and with the short TRACON session in Study 1,

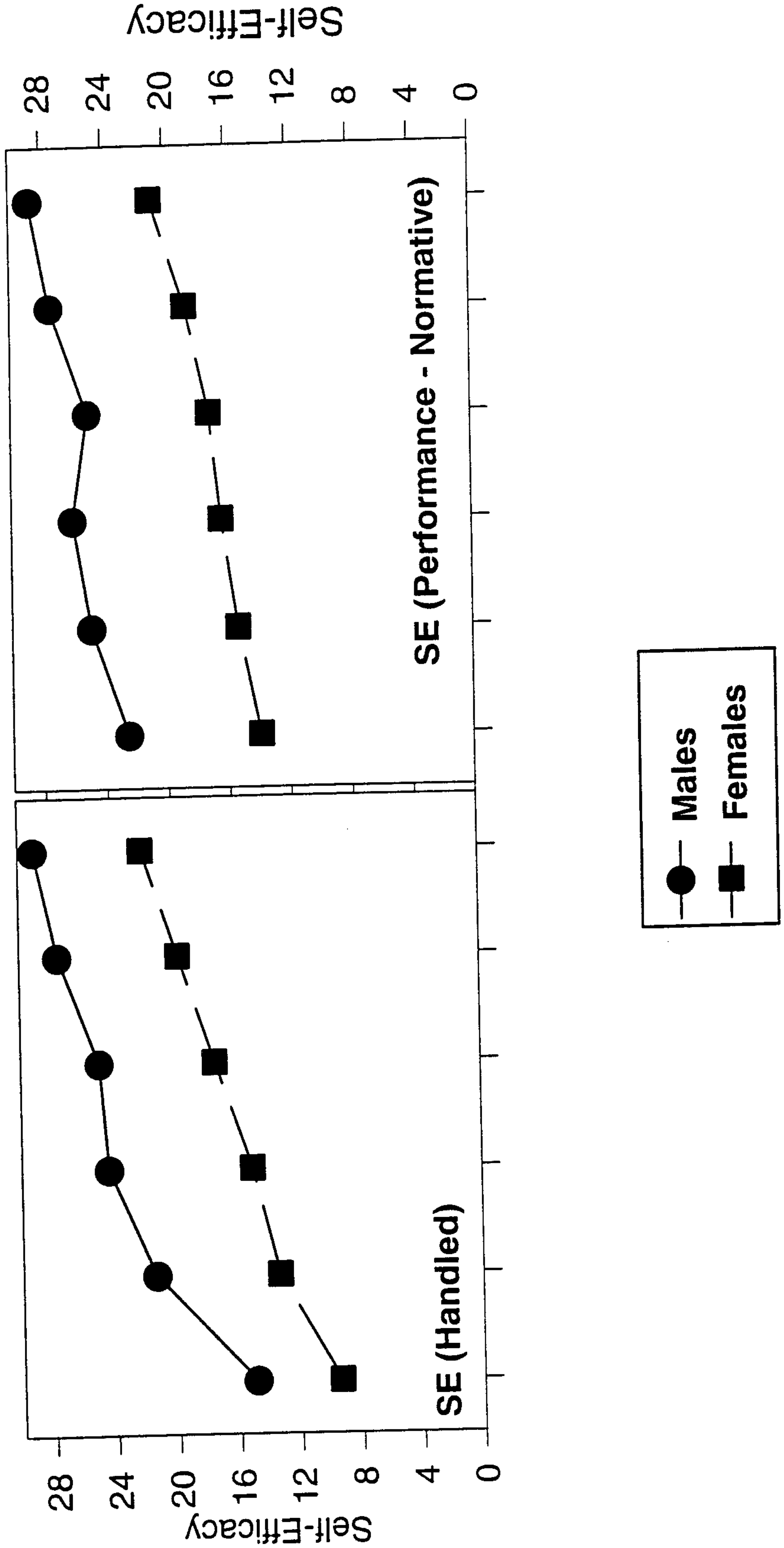


Figure 19. Mean Self-Efficacy expectation composite scores over TRACON practice, by gender. Measures administered *after* each TRACON session.

Table 5. Hierarchical Multiple Regression/Correlation Results for Study 3. Variance accounted for.

	Criterion					
	Overall Performance		Overflights		Arrivals	
Predictor	Session 1	Session 6	Session 1	Session 6	Session 1	Session 6
Ability	38.4%**	44.0%**	33.1%**	22.6%**	31.7%**	44.7%**
Motivational Skills	1.6	1.1	1.2	0.0	0.0	1.3
Self-estimates of ability/ Self-concept	1.5	1.9	1.0	1.4	3.8	1.7
Interests (Real)	0.0	0.1	0.1	0.4	0.4	0.5
Negative Motivation/ Positive Motivation	3.8	6.1**	3.7	9.5**	2.9	2.0
Self-Efficacy (pre-task)	0.0	0.8	0.1	0.0	0.8	1.1
Gender	1.5	2.9*	0.0	0.0	4.4*	6.1**
Total Variance Accounted For	46.7	56.9	39.0	35.1	44.0	57.5

Note. * $p < .05$; ** $p < .01$.

cognitive and perceptual speed abilities accounted for the largest amount of individual differences variance in overall TRACON task performance, as well as performance in the overflight and arrival components of the TRACON task, *across extensive practice*. Of the remaining predictor measures, only proximal Negative and Positive Motivational thoughts consistently provided incremental validity -- which was more pronounced in the overflight component than the arrival task component. Finally, as with previous studies, gender accounted for a significant incremental amount of variance in the arrival component of task performance, but accounted for virtually no incremental variance in the overflight component.

Perceptual Speed (complex) Ability Followup

We also subjected the composite made up of the Dial Reading Test and the Directional Headings Test to cross-correlation analyses with non-ability predictors of performance in TRACON. Similar to the types of correlations found in Study 1 (with only the Dial Reading Test), the composite correlated significantly with several non-ability measures, as follows: Interests (Realistic) - r = .35, Self-estimate of ability (spatial-mechanical- and science- self-concept) - r = .34, and Self-efficacy (performance) across practice (r's = .37, .33, .38, .39, .36, and .31) for Sessions 1 - 6 respectively. Of considerable importance were the correlations between the composite and initial Negative Motivation and Positive Motivation thoughts. Consistent with our interpretation of the non-ability influences on performance in the Dial Reading and Directional Headings Tests, the composite correlated negatively with Negative Motivation scores (r = -.31), indicating that low scores on the composite were associated with greater reporting of negative motivational thoughts prior to task practice. In addition, the composite showed essentially a zero correlation with Positive Motivational Thoughts prior to Session 1 (r = .07), *even though Negative and Positive Motivation Thoughts were highly negatively correlated* (i.e., r = -.58). Also, a further result that was consistent with our speculation about the tests, the composite correlated significantly with a self-report measure of Stress-Resistance Self-Concept (r = .28) -- the Perceptual Speed (complex) composite was the only objective ability test composite that significantly correlated with this non-ability measure.

Finally, we subjected the Perceptual Speed (complex) composite to a gender analysis. Even though the constituent tests have some surface-level spatial content (which would lead to an expectation of higher mean scores for males), only a marginally significant difference was found for gender (M_{females} = -.18, M_{males} = .22, $t(91)$ = 1.96, p = .05).

Discussion

The increased time available in Study 3 for pre-task ability testing and assessment of individual differences in a variety of non-ability predictors of performance, and for greater criterion task practice time, provided for a wider assessment of the predictor construct space, and for the study of performance on a more complex skill-learning task than was possible in the previous laboratory and field studies. In many ways, the TRACON task is more representative of the kind of real-world tasks for which performance under stress is observed.

The task is continuous (over a watch of 30 min), can involve unpredictable events, and has substantial information processing stress (by virtue of aircraft pilots repeated requests for permissions over the headset, and when separation conflicts are in progress and warning alerts are broadcast visually and auditorally). The only salient aspects of the task that are not characteristic of the real-world is that no actual loss of human life occurs when mistakes are made, and no direct physical threat is imposed on the individual operator. Nonetheless we have observed behaviors in the laboratory that are consistent with the inference that the task is stressful. Incidences of subjects in tears, cursing, becoming agitated with the equipment (e.g., pounding the keyboard or trackball), have been observed over the course of our experiments with TRACON. The observed Negative Motivation Thought occurrences between task sessions are consistent with this inference as well -- some subjects show worry and apprehension prior to the TRACON sessions, and report negative affective reactions at the conclusion of TRACON sessions. Although there is controversy in the field about whether the job of air traffic controller is of greater stress than other jobs, there is virtually no doubt that the task is stressful during the skill acquisition phase.

The results from Study 3 indicated that many ability and non-ability predictors provide significant correlations with individual differences in performance on TRACON over practice. The non-ability predictors range from distal vocational interests to academic and non-academic self-concept, motivational skills to proximal reports of negative and positive motivational thoughts and self-efficacy expectations. As with the previous two studies, though, when objective ability measures are first entered into a multiple regression equation for predicting task performance, little incremental validity is found for the non-ability predictors. In this study, only incidents of proximal negative and positive motivational thoughts added significant incremental prediction to the equation. That is, common variance between most of the non-ability predictors and TRACON performance is accounted for by several objective ability composite measures. Most prominent among these ability predictors were the two tests identified as Perceptual Speed (complex), namely the Dial Reading Test and the Directional Headings Test. As in Study 1, such variables shared considerable common variance with the important non-ability predictors of performance. Such results again support the inference that both ability and non-ability influences of individual differences in performance under stressful conditions can be well accounted for by a set of objective tests with characteristics similar to the Dial Reading and Directional Headings Tests.

VIII. General Discussion and Conclusions

The first fundamental perspective that was undertaken with this series of studies was that theory-based selection of cognitive ability and non-ability predictors of performance could result in significant and substantial prediction of individual differences in task performance under a variety of conditions which included information-processing task stressors. Over the course of conducting three studies, we found that several major cognitive and perceptual speed ability composites were instrumental in predicting individual differences in performance under both more and less stressful task-load conditions. In addition, we validated several distal self-report measures (e.g., self-concept, self-ratings of ability, motivational skills, vocational interests) and proximal self-report measures (self-efficacy expectations and frequency of negative and positive motivational thoughts) for predicting performance under various task conditions. In the aggregate, these measures were able to predict roughly 30% of the individual-differences variance in performance on the Target/Threat Identification Task, and roughly 50% of the individual differences variance in performance on the TRACON air traffic controller simulation -- these were significant and substantial accounts of individual differences in performance. Similar to an earlier evaluation of the existence of a time-sharing ability (that is, a special ability to perform more than one task at a time -- see Ackerman, Schneider, & Wickens, 1982, 1984), where constituent task competencies account for much greater variance than a specific time-sharing ability; we found that identified predictors of performance did not change their influence across stressed and non-stressed conditions.

After examination of the influence of individual predictor variables (or families of predictor variables) on performance, we conducted a series of hierarchical multiple regression/correlation analyses, in order to assess the incremental validity of the non-ability measures in predicting individual differences in task performance. In general, two major findings resulted from these analyses: (1) Cognitive and Perceptual Speed abilities accounted for the majority of explained variance under various task conditions; and (2) Non-ability predictors of performance, while significantly related to task performance, rarely provided significant incremental validity in predicting performance, *after* cognitive and perceptual abilities had been entered into the regression equation. That is, individual differences variance in performance, accounted for by non-ability predictors, was shared variance with the sets of cognitive and perceptual speed ability measures.

More in-depth analyses of the ability and non-ability predictors indicated that two tests were mainly responsible for capturing the common variance among predictors, namely a test originally validated for WWII aircrew selection (Dial Reading Test) and a test that was originally validated for FAA air traffic controller selection (Directional Headings Test). These two tests, while nominally identified as representing Perceptual Speed, Spatial, and Numerical abilities, also shared significant variance with distal non-ability measures of vocational interests, self-ratings of ability and self-concept, as well as with proximal non-ability measures of negative motivational thoughts and self-efficacy expectations. Closer examination of the content of the Dial Reading and Directional Headings tests suggested that these tests are highly susceptible to frustrations and emotional responses that have similar characteristics to task performance under stress-loaded conditions.

Although there has been an extensive literature concerning the Dial Reading and Directional Headings tests to date, investigators have not examined the non-ability determinants of performance on these tests, *even though these tests have demonstrated the highest validity in large batteries of ability tests for predicting job performance under stress.* Our results strongly suggest that future investigations should be conducted to broaden the construct space underlying these measures, for example, by developing a battery of tests that have similar information-processing characteristics. With a larger battery of these tests, it may be possible to independently account for the ability and non-ability determinants of performance on such measures. Such a result could have important implications for implementation of new assessment instruments for selection of applicants for jobs that involve substantial stress exposure.

Whenever common variance is detected among several assessment instruments, as was the case for the Dial Reading and Directional Headings Tests on the one hand, and the non-ability measures on the other hand -- it is a judgement call as to which set of measures should be implemented in practice. The advantages of the kinds of tests represented by the Dial Reading and Directional Headings tests for an operational environment are clear -- these are objective tests (as opposed to self-report, as was the case for the non-ability measures studied in this project). For implementation purposes, the objective "ability" tests are substantially less susceptible to coaching or to response biases.

In the course of this project, we had hoped to be able to identify aspects of individual differences in responses to stressful task demands that might be conducive to remediation. Examination of both the ability and non-ability predictors of performance under stress-loaded conditions suggests that performance deficits under stressful conditions are influenced by the same self-regulatory deficiencies that adversely affect performance in complex tasks under normal (or non-stressed) conditions. Most prominent among these deficiencies is Emotion Control -- a skill that some individuals have developed that allows them to maintain task-relevant attention in the face of errors and goal-attainment frustrations. Although we have previously demonstrated that it is indeed possible to develop effective training methods for developing emotion control skills (see Kanfer & Ackerman, 1990), it is currently impossible to determine whether such training might have transfer effects to conditions of ability testing (as in the Dial Reading and Directional Headings tests). Speculation at this point is that training that focuses on remediating the adverse emotion control reactions to novel or stressful task situations may indeed have the desired effect of developing additional resistance-to-stress under performance demands. We recommend that, subsequent to mapping out the construct space underlying the Dial Reading and Directional Headings tests, a study be carried out that tests this inference. In this way, it may be possible to delineate whether emotion-control reactions are malleable (and thus should be handled in a training context) or are relatively resistant to training (and thus should be handled in a selection context).

IX. References

Ackerman, P. L. (1986). Individual differences in information processing: An investigation of intellectual abilities and task performance during practice. *Intelligence, 10*, 101-139.

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117*, 288-318.

Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinants. *Journal of Applied Psychology, 77*, 598-614.

Ackerman, P. L., & Goff, M. (1994) Typical intellectual engagement and personality: Reply to Rocklin (1994). *Journal of Educational Psychology, 86*, 150-153.

Ackerman, P. L., & Kanfer, R. (1993). Integrating laboratory and field study for improving selection: Development of a battery for predicting air traffic controller success. *Journal of Applied Psychology, 78*, 413-432.

Ackerman, P. L., Schneider, W., & Wickens, C. D. (1982). *Individual differences in time-sharing ability: A critical review and analysis* (ONR Report #8102). Champaign, IL: University of Illinois, Human Attention Research Laboratory, (ERIC Document Reproduction Service No. ED 221 550).

Ackerman, P. L., Schneider, W., & Wickens, C. D. (1984). Deciding the existence of a time-sharing ability: A combined theoretical and methodological approach. *Human Factors, 26*, 71-84.

Alexander, J. R., Ammerman, H. L., Fairhurst, W. S., Hostetler, C. M., & Jones, G. W. (1990). FAA air traffic control operations concepts volume VI: ARTCC/Host en route controllers. Washington, D.C.: FAA report no. DOT/FAA/AP-87-01.

Ammerman, H. L., Becker, E. S., Jones, G. W., Tobey, W. K., & Phillips, M. K. (1987). *FAA air traffic control operations concepts volume I: ATC background and analysis methodology*. Washington D.C.: FAA report no. DOT/FAA/AP-87-01.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369-406.

Anderson, J. R. (1985). *Cognitive psychology and its implications (2nd ed.)*. New York: W. H. Freeman and Company.

Bandura, A. (1986). Self-regulation of motivation and action through goal systems. In V. Hamilton, G H., Bower, & N. H. Fryda (Eds.), *Cognition, motivation, and affect: A cognitive science view* (pp. 37-61). Dordrecht, Holland: Martinus Nijhoff.

Buckley, E. P., DeBaryshe, B. D., Hitchner, N., & Kohn, P. (1983). *Methods and measurements in real-time air traffic control system simulation* (DOT/FAA/CT-83/26). Atlantic City Airport, NJ: U.S. Department of Transportation, Federal Aviation Administration.

Campbell, J. P., & Pritchard, R. D. (1976). Motivation theory in Industrial and Organizational Psychology. In M. D. Dunnette (Ed.), *Handbook of Industrial and Organizational Psychology*, 63-130).

Cobb, B. B., & Mathews, J. J. (1972). *A proposed new test for aptitude screening of air traffic controller applicants* (FAA-AM-72-18). Washington, DC: U.S. Department of Transportation, Federal Aviation Administration.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five Factor Inventory (NEO-FFI): Professional Manual.* Odessa, FL: Psychological Assessment Resources.

Dawis, R. V., & Lofquist, L. H. (1984). *A psychological theory of work adjustment: An individual-differences model and its applications.* Minneapolis, MN: University of Minnesota Press.

Dickman, S. J. (1990). Functional and dysfunctional impulsivity: Personality and cognitive correlates. *Journal of Personality and Social Psychology, 58,* 95-102.

Donchin, E., Fabiani, M., Sanders, A. (Eds.) (1989). The learning strategies program: An Examination of the Strategies in Skill Acquisition. *Acta Psychologica, 70,* 1-309.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Kit of factor-referenced cognitive tests.* Princeton, NJ: Educational Testing Service.

Fisk, A. D., Ackerman, P. L., & Schneider, W. (1987). Automatic and controlled processing theory and its application to human factors problems. In P. A. Hancock (Ed.), *Human Factors Psychology*, North-Holland.

Fitts, P. & Posner, M. I. (1967). *Human performance.* Belmont, CA: Brooks/Cole.

Frederiksen, J. R., & White, B. Y. (1989). An approach to training based upon principled task decomposition. *Acta Psychologica, 70*, 89-146.

Goff, M., & Ackerman, P. L. (1992). Personality-intelligence relations: Assessing typical intellectual engagement. *Journal of Educational Psychology, 84,* 537-552.

Guilford, J. P. (1982). Cognitive psychology's ambiguities: Some suggested remedies. *Psychological Review, 89*, 48-59.

Guilford, J. P., & Lacey, J. I. (Eds.). (1947). *Army air forces aviation psychology program research reports: Printed classification tests.* Report No. 5. Washington, DC: U. S. Government Printing Office.

Helmreich, R. L., & Spence, J. T. (1978). Work and family orientation questionnaire: An objective instrument to assess components of achievement motivation and attitudes toward family and career. *Psychological Documents*, 1677. American Psychological Association.

Holland, J. L. (1959). A theory of vocational choice. *Journal of Counseling Psychology, 6,* 35-45.

Holland, J. L. (1985). *Making vocational choices: A theory of vocational personalities and work environments* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*, 253-270.

Humphreys, M. S., & Revelle, W. (1984). Personality, motivation, and performance: A theory of the relationship between individual differences and information processing. *Psychological Review, 91*, 153-184.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

Kanfer, F. H. (1977). The many faces of self-control, or behavior modification changes its focus. In R. B. Stuart (Ed.), *Behavioral self-management*. New York: Brunner/Mazel.

Kanfer, R. (1987). Task-specific motivation: An integrative approach to issues of measurement, mechanisms, processes, and determinants. *Journal of Social and Clinical Psychology, 5*, 237-264.

Kanfer, R. (1991). Motivation theory and Industrial/Organizational psychology. In M. D. Dunnette and L. Hough (Eds.), *Handbook of industrial and organizational psychology. Volume 1. Theory in industrial and organizational psychology*. Palo Alto, CA: Consulting Psychologists Press.

Kanfer, R. & Ackerman, P. L. (1990). Ability and metacognitive determinants of skill acquisition and transfer. Final Report to Air Force office of Scientific Research. Minneapolis: University of Minnesota.

Kanfer, R. (July, 1994). *Motivation, self-regulation, and self-efficacy: a skills perspective.* Presentation at the 1994 annual convention of the American Psychological Society. Washington, DC.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology Monograph, 74*, 657-690.

Kleinbeck, U. (1987). The effects of motivation on job performance. In F. Halisch and J. Kuhl (Eds.), *Motivation, intention, and volition*, 261-271. New York: Springer - Verlag.

Kluwe, R. H., & Friedrichsen, G. (1985). Mechanisms of control and regulation in problem solving. In J. Kuhl & J. Beckmann (Eds.) *Action control:* From cognition to behavior, 183-218. New York: Springer - Verlag.

Kuhl, J. (1985). Volitional mediators of cognition - behavior consistency: Self-regulatory processes and action versus state orientation. In J. Kuhl and J. Beckmann (Eds.), *Action control: From cognition to behavior* (pp. 101-128). New York: Springer-Verlag.

Lamb, R. R., & Prediger, D. J. (1981). *The Unisex Edition of the ACT Interest Inventory.* Iowa City, IA: American College Testing.

Landon, T. A. (1991). *Job Performance for the en route ATCS: A review with applications for ATCS selection.* Unpublished report to the Minnesota Air Traffic Controller Training Center.

Locke, E. A. (1968). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance, 3*, 157-189.

Lohman, D. F. (1986). The effect of speed-accuracy tradeoff on sex differences in mental rotation. *Perception & Psychophysics, 39,* 427-436.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence*, *7*, 107-127.

Means, B., Mumaw, R., Roth, C., Schlager, M., McWilliams, E., Gagné, E., Rice, V., Rosenthal, D. & Heon, S. (1988). *ATC training analysis study: design of the next-generation ATC training system.* Technical report OPM/342-036. Washington, DC: FAA, Office of training and higher education.

Mitchell, T. R. (1982). Expectancy-value models in organizational psychology. In N. T. Feather (Ed.), *Expectations and actions: Expectancy-Value Models in psychology,* 293-312. Hillsdale, NJ: Erlbaum.

Naylor, J. C., Pritchard, R. D., & Ilgen, D. R. (1980). *A theory of behavior in organizations.* New York: Academic Press.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition*. Hillsdale, NJ: Erlbaum.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, *7*, 44-64.

Schneider, W., & Fisk, A. D. (1982). Dual task automatic and control processing, can it occur without resource cost? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 261-278.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127-190.

Snow, R. E., Kyllonen, P. C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 47-103). Hillsdale, NJ: Erlbaum.

Spielberger, C. D. (1983). *The State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Technical Supplement to the Counselor's Manual for the Armed Services Vocational Aptitude Battery Form-14 (1986). Chicago, IL: U.S. Military Entrance Processing Command.

Tellegen, A. (1982). *Brief manual for the Multidimensional Personality Questionnaire (MPQ)*. Copyright 1982 by Auke Tellegen.

Vernon, P. E. (1961). *The structure of human abilities*. New York: Wiley.

Vickers, R. R., Jr. (1991). *Stress reactivity: Five-factor representation of a psychobiological typology* (Report No. 91-26). San Diego, CA: Naval Health Research Center.

Vroom, V. H. (1960). *Some personality determinants of the effects of participation*. Englewood Cliffs, NJ: Prentice-Hall.

Werdelin, I., & Stjernberg, G. (1969). On the nature of the perceptual speed factor. *Scandinavian Journal of Psychology*, *10*, 185-192.

Wickens, C. D. (1984). *Engineering psychology and human performance*. Columbus, OH: Charles E. Merrill Publishing.

X. Appendix A -- Target/Threat Identification Task

Main Display. The task is made up of one main screen, and three informational feedback screens. The main screen displays a simulated battle display (see Figure A1), with stationary ships (shown as silhouettes), each surrounded by two concentric circles, a red (or danger circle) on the inside, and a white (firing circle) on the outside. "Targets" are also displayed (which appear as yellow triangles) -- each is in motion approaching one of the ships. A box in the upper left portion of the screen displays a list of identifiers (arbitrary nouns, such as animal names), and for each identifier, an indication of "Friend" or "Enemy" status. The trainee's task is to 'hook' each of the targets, by using a mouse to position a cursor over the target icon and pressing the left mouse button (which brings up the display of the identifier for that plane). The identifier stays on the screen until the trainee confirms the identification (Friend or Enemy) by pressing corresponding keys on the computer keyboard. Once the target is identified, the icon changes color. Each target is to be identified during the course of the task trial. As the trial progresses, each target moves toward a ship on the display. When the target reaches the outer circle (firing circle), the subject may 'fire on' the target, by moving the mouse cursor over the target icon, and pressing the right mouse button (which will destroy the target, and remove it from the screen). All Enemy targets are to be destroyed before they cross the red ('danger') circle. Friendly targets are to be allowed to pass the red circle. When targets reach the red circle, they either disappear (friendly targets) from the display, or fire on the ships (resulting in an explosion display), and then disappear from the screen. The trial ends when all planes have either crossed the respective red circles, or have been destroyed.

Feedback Displays. At the end of each trial, three feedback displays are presented (see Figure A2). The first display presents trial results with a four-fold figure, that presents information in a signal-detection framework. The four categories listed [and the points assigned for each category) are: friendly targets passed (Correct Rejection = +750 points), friendly targets destroyed (False Alarm = -1500 points), enemy targets passed (Miss = -2500 points), and enemy targets destroyed (Hit = +750 points). The display also shows the number of redundant queries (targets that are queried more than once = -50 points), and total score for the trial.

The second feedback display presents a text summary of performance, that includes the following information:

1. Number of targets correctly identified (first query only)
2. Number of redundant queries (targets queried more than once)
3. Number of targets not queried
4. Percent correctly identified (all queries)

The final feedback display shows a bar graph of total score for the current trial, and for previous trials (up to five previous trials).

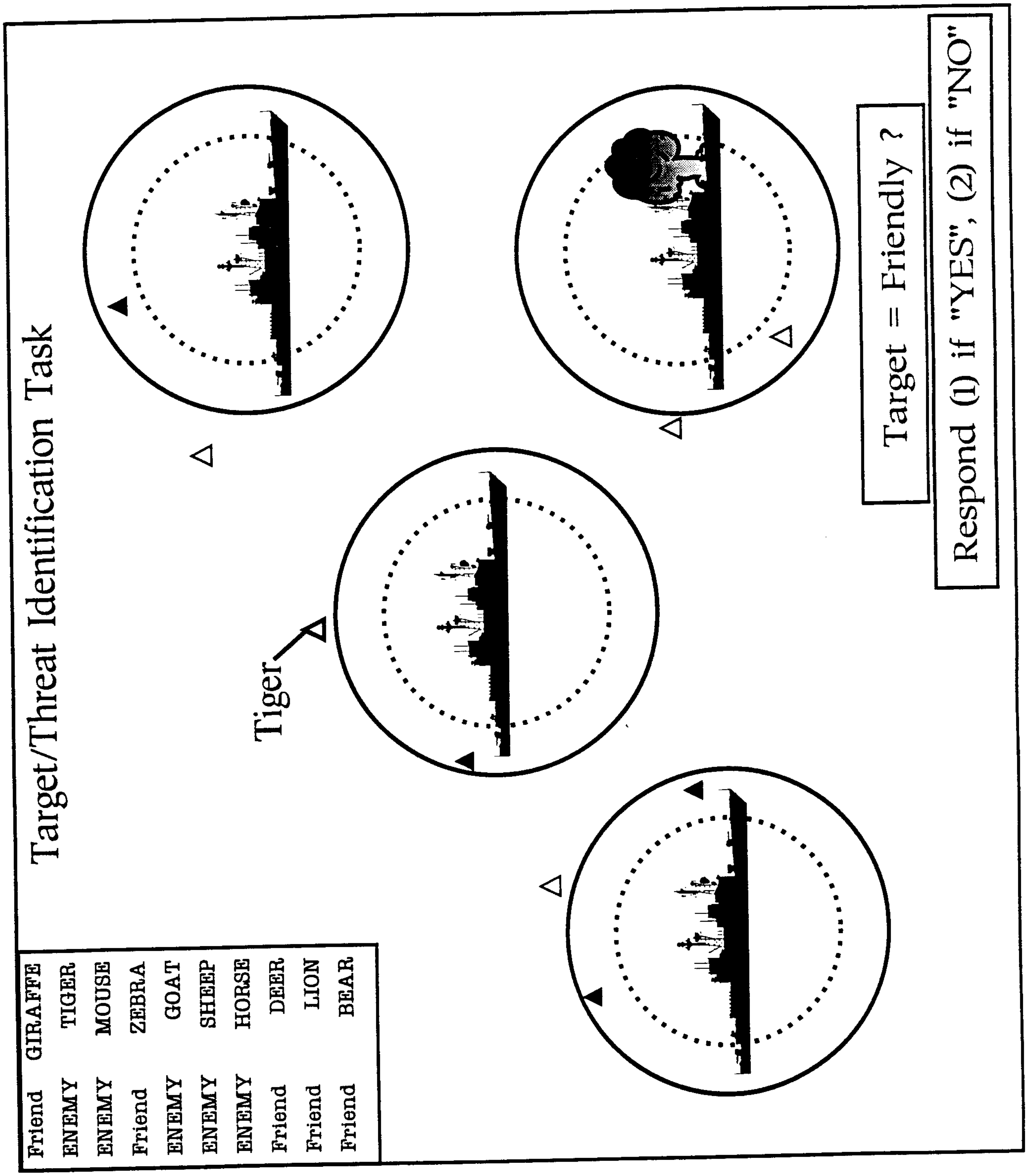


Figure A1. Target/Threat Identification Task main stimulus display. See text for details.

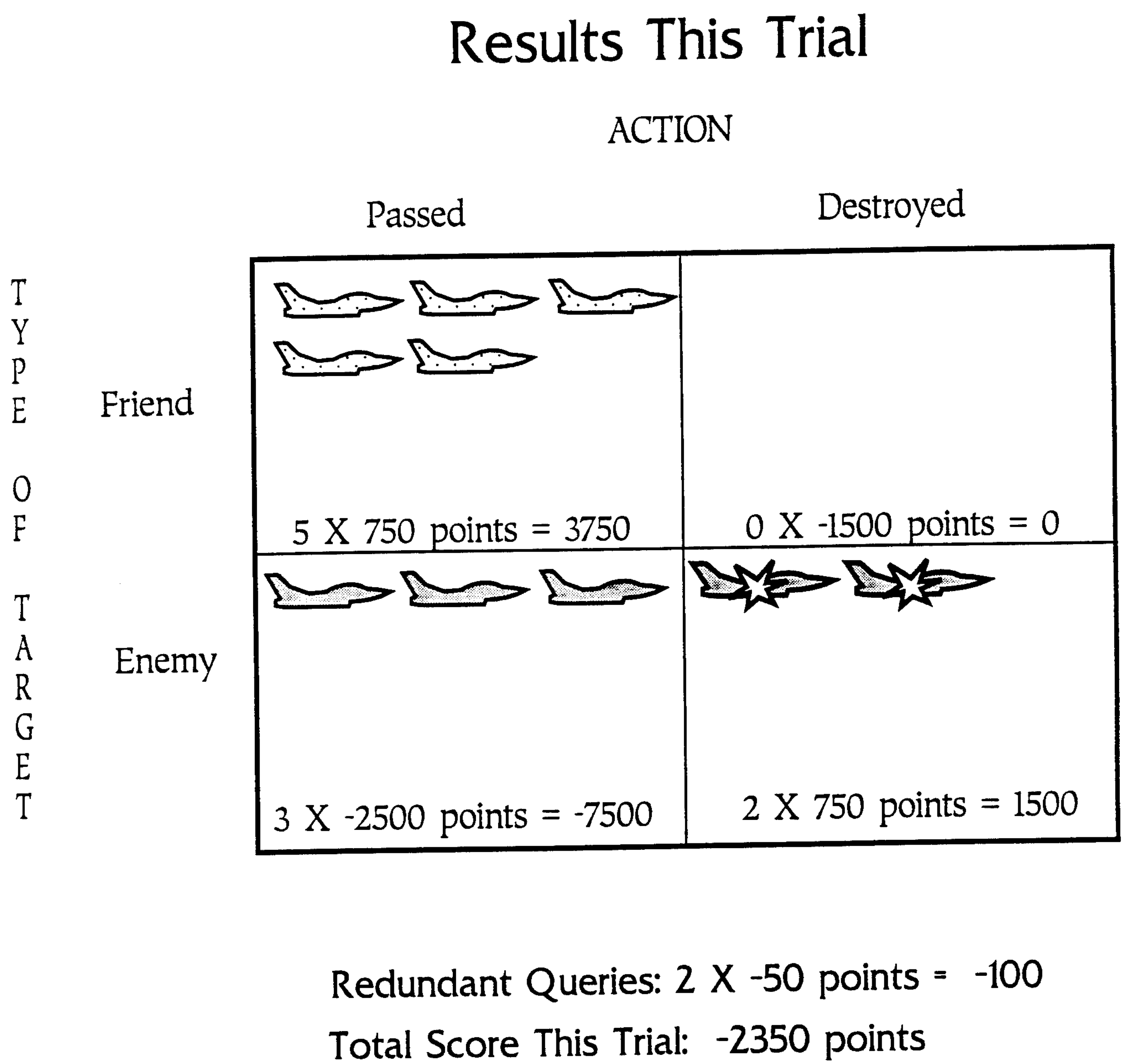


Figure A2. Target/Threat Identification Task first feedback display. See text for details.

Trial description Trials for the task were created and pretested to be roughly equivalent in difficulty. Each trial contained four ships and 10 targets (5 friendly and 5 enemy). Each trial was constrained to have a total duration of 24 seconds. The trial begins with targets at different distances from the ships, that require them to travel 18-24 sec to reach the inner circle around the ships. Specifically, 2 targets take 18 sec, 3 take 20 sec, 3 take 22 sec and 2 take 24 sec, respectively to reach the ships. There is a constant 13 sec interval between the time targets traverse the outer to the inner circles around the ships.

CM Conditions. For consistent mapping trials, the identifiers and friend/enemy status are constant throughout the trial sequence. As such, trainees that form an associative memory representation of the mappings, no longer need to look up the friend/enemy status on the display. To assess the memory representation of this information, a memory test is administered at the end of the CM trials. The memory test randomly presents the list of target identifiers and requires the trainee to indicate friend or enemy status, in the absence of the look up table.

VM Conditions. For varied mapping trials, the identifiers and friend/enemy status are randomly paired at the beginning of each trial. As such, the trainees cannot memorize the identifier-status associations across trials.

Dual-Task Conditions. During the dual-task conditions, an audio tape was presented over the public address system. The audio tape was an actual radio recording of controller-pilot conversations at a local FAA tower facility. The trainees were instructed to listen to the tape while they performed the target/threat task, and at the end of practice, they would be instructed to recall as many plane identifiers that were mentioned on the tape. At the end of task practice, the trainees were given an open-ended recall task for these identifiers.

Performance measurement. Performance on the Target/Threat task was multiply assessed. In addition to the raw measures for each trial, sensitivity was calculated with a non-parametric measure, called A' (see Wickens, 1984). Total score was calculated on the basis of point assignments for each type of activity. Reaction time was calculated as a mean identification time (elapsed time from target 'hook' to key-press indicating friend/enemy). For the CM conditions, memory test performance was assessed as the total number of correct responses. For the dual-task conditions, the number of correct plane identifier recalls was calculated.

XI. Appendix B -- Terminal Radar Approach Control (TRACON) Simulation Task

The task used for this research is a uniquely modified Professional version (V1.52) of TRACON simulation software, developed by Wesson International. Versions of this program have been, and are, in use in several locations in this country (including the FAA, NASA, DoD, and several colleges and university airway sciences programs) for training of air traffic controllers. Modifications for the current instantiation of the program allowed for the collection of a variety of data, described in more detail below.

Task analysis is a critical component of empirical investigation with any task, but especially with a complex task such as TRACON (or the task used by Donchin, et al., 1989; see also Frederiksen & White, 1989). In fact, one of the salient reasons for adoption of the TRACON task here is that there has been extensive historical work on task analysis for the full range of Air Traffic Controller tasks, including the subset implemented in TRACON. In particular, a five-volume set of task analysis materials has been prepared by the FAA (e.g., Alexander, et al., 1990; Ammerman, et al. 1987). Additional task analyses have been conducted by the U.S. Air Force, and more recently by HumRRO (Means et al., 1988). All of these have been carefully studied in an effort to derive the critical components of the simulation task adopted for empirical evaluation (see, e.g., Landon, 1991).

Descriptions of the TRACON simulations are provided in Ackerman (1992) and Ackerman and Kanfer (1993). The following discussion abstracts that material, with deviations in procedures specifically noted. The task requires that trainees learn a set of rules for air traffic control, including reading flight strips, declarative knowledge about radar beacons, airport locations, airport tower handoff procedures, en-route center handoff procedures, plane separation rules and procedures, monitoring strategies, and strategies for sequencing planes for maximum efficient and safe sector traversal. In addition, trainees are required to acquire human-computer interface skills: including issuing trackball-based commands, menu retrieval, keyboard operations, and integration between visual and auditory information channels. Note, however, that the simulator task represents a substantial reduction of rules and operational demands in comparison to the real-world job of ATCs.

Display. TRACON presents the trainee with a simulated color radar screen, depicting a region of airspace, Very High Frequency Omnidirectional Range Stations (VOR), airports, sector boundaries, and range rings. Planes are identified by an icon on the radar scope, with a data tag that indicates plane identification and altitude information. In addition, two sets of "Flight Strips" are presented at the right side of the display, a "Pending" and an "Active" set. Each flight strip contains information about a particular flight, including identification information, plane type, requested speed and altitude, and sector entry and exit destination information (See Figure B1). Finally, at the bottom of the screen is a "Communications Box" -- which shows commands issued to planes (and responses by the pilots), along with the controller's "score" for the current simulation. When planes were about to enter the trainee's sector (at a boundary or on the runway of an airport) this information was announced over the headset. No flight was allowed to cross the sector boundary or take off from an airport without explicit authorization by the trainee.

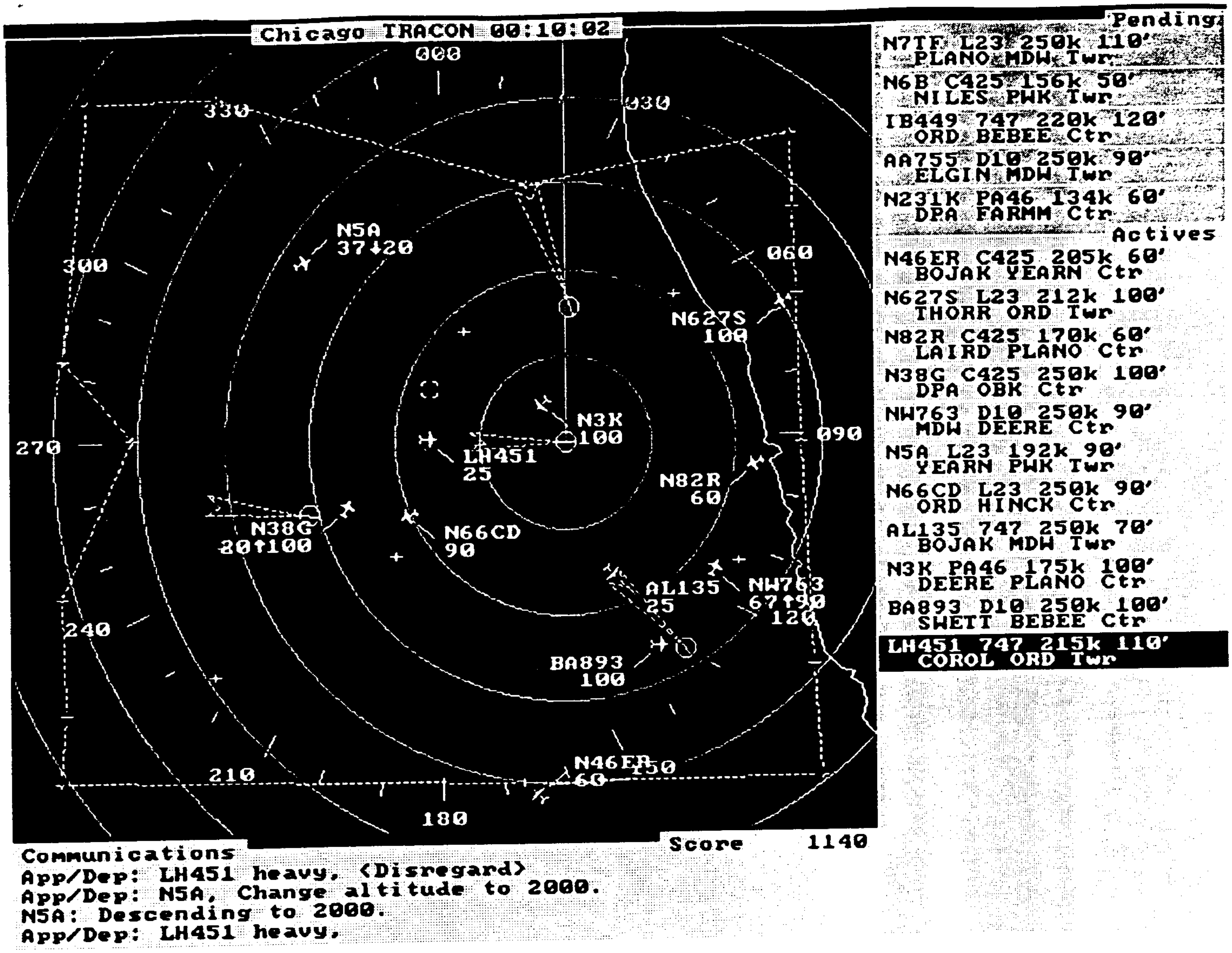

Chicago TRACON 00:10:02
Pending
N7TF L23 250k 110'
PLANO MDW Twr
N6B C425 156k 50'
NILES PWK Twr
IB449 747 220k 120'
ORD BEBEE Ctr
AA755 D10 250k 90'
ELGIN MDW Twr
N231K PA46 134k 60'
DPA FARMM Ctr
Actives
N46ER C425 205k 60'
BOJAK YEARN Ctr
N627S L23 212k 100'
THORR ORD Twr
N82R C425 170k 60'
LAIRD PLANO Ctr
N38G C425 250k 100'
DPA OBK Ctr
NW763 D10 250k 90'
MDW DEERE Ctr
N5A L23 192k 90'
YEARN PWK Twr
N66CD L23 250k 90'
ORD HINCK Ctr
AL135 747 250k 70'
BOJAK MDW Twr
N3K PA46 175k 100'
DEERE PLANO Ctr
BA893 D10 250k 100'
SWETT BEBEE Ctr
LH451 747 215k 110'
COROL ORD Twr
Score 1140
Communications
App/Dep: LH451 heavy, <Disregard>
App/Dep: N5A, Change altitude to 2000.
N5A: Descending to 2000.
App/Dep: LH451 heavy,

Figure B1 (Preceding Page). Static copy of TRACON® screen. There are three major components to the display. The right hand side of the screen shows Pending (not under control) and Active (under control) flight strips. Each flight strip lists (a) plane identifier, (b) plane type, (c) requested speed, (d) requested altitude, (e) Radar fix of sector entry, (f) Radar fix of sector exit (including Tower or Center). The lower part of the screen shows a communications box that gives a printout of the current (and last few) commands issued by the subject, and the responses from pilots or other controllers. The main part of the screen shows a radar representation of the Chicago sector. Planes are represented by a plane icon, and a data tag (which gives the identifier, the altitude, and an indication of current changes in altitude). The sector is bounded by the irregular dotted polygon describing a perimeter. Radar fixes are shown as small (+) figures on the radar screen. Airports are shown with approach cones, and a circle indicating the facility proper. A continuous radar sweep is shown (updating at 12 o'clock, every 5 sec). Range rings are also displayed, indicating 5 mile distances.

Task controls and knowledge of results. Trainees interacted with the TRACON simulation in several ways. A trackball was used for the majority of input activities, although the keyboard was also used alone, or in conjunction with the trackball (the trackball represents a change in input device from the previous study, where a mouse was used). For each plane command, a menu of command choices was displayed on the screen.

Knowledge of results was provided visually (by text in the communications box) and aurally with a read-back by the pilot or other controller (using digitized speech). In addition, planes followed (as nearly as possible) the commands issued by the trainee. Turn, altitude change, and speed change commands were processed by the computer, and were carried out in accordance with the limitations imposed by each aircraft type.

When errors occurred (e.g., separation conflicts, near misses, crashes, missed approaches, handoff errors), additional information was presented to the trainee. In each of these cases, an alert circle around the plane(s) in question was presented on the screen, and a series of tones were presented over the headset.

Trial description Trials for the task were created and pretested to be roughly equivalent in difficulty. Each trial contained planes that were divided into three basic categories (Overflights, Departures, and Arrivals). Overflights were planes that entered and exited the trainee's airspace at cruising altitudes. Trainees were required to acknowledge these airplanes as they approached a boundary VOR fix, monitor progress through the sector, and handoff to a "Center" controller. Departures were planes that originated at one of the four airports, climbed to a cruising altitude and were handed off to a "Center" controller. Trainees were required to release departures from airports, evaluate and remediate potential conflicts as the planes climbed to a cruising altitude and turned to intercept their intended flight paths, and then handoff planes to the appropriate Center controller. Arrivals entered the trainee's airspace from one of the boundary VOR fixes, and had to be landed at a designated airport. Trainees were required to direct arrivals onto an appropriate heading and altitude to provide an acceptable handoff to the appropriate Tower controller, then these planes would land. For all flights, the trainee was required to maintain legal separation (at least 1000 ft. [304.8 m] in altitude, or 3 mi [5.56 km] horizontally).

Each trial was comprised of 16 overflights and departures (with roughly equal frequency), and 12 arrivals. The planes requested entry to the airspace at irregular intervals that were constrained to require the trainee to be always occupied with at least one active target. The trials were also constrained so that perfect performance (handling all 28 planes successfully) was just beyond the skill level achieved by subject matter experts. Each trial was concluded in 30 min.

A successful "handle" of a flight was the appropriate accomplishment of the respective flight plan. That is, for a departure or an overflight, the accomplishment was a successful handoff to the appropriate Center controller. For a landing, the accomplishment was the successful landing of the airplane.

Performance measurement. As with a previous investigation (Ackerman, 1992), overall performance was computed as the sum of all flights accepted into the sector that had a final disposition within the simulation time (minus any planes that were incorrectly disposed of -- e.g., crashes, not-handed-off, vectored off the radar screen). This measure is concordant with measures derived from the examination of the criterion space for FAA ATC simulation research (e.g., see Buckley, Debaryshe, Hitchner, & Kohn, 1983), and has acceptable reliability.